Christopher Probert
cprobert@stanford.edu
CS231A Project Report

Segmentation and descriptors for cellular immunofluoresence images

1: Introduction

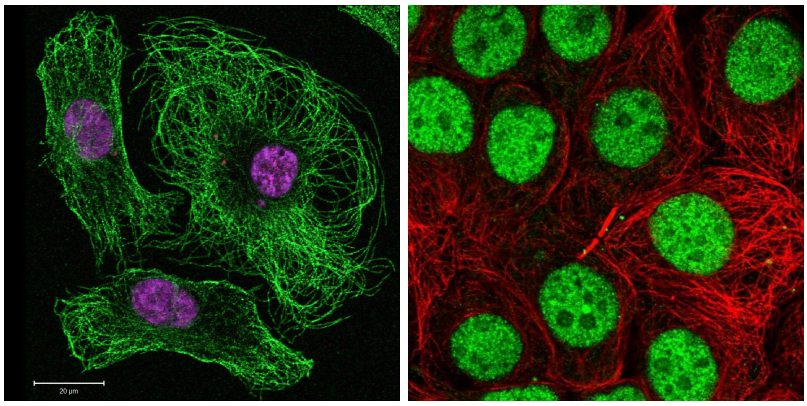


Figure 1: Example immunoflouresence images of human cancer cells. Different colors correspond to different fluorescent antibodies, which show the location of different molecules.

Some of the most important open questions in studying cancer are those of intra-tumor heterogeneity: how do the mixtures and spatial distributions of different cell types vary between tumors, and what are the implications on disease? Some of the most promising approaches for studying these problems are imaging based, such as immunofluorescence. Immunofluorescent imaging uses specific molecular tags (called antibodies) to identify molecules within a cell. Each antibody is covalently attached to a fluorescent marker molecule (fluorophore) that is excited by a specific wavelength (channel) of light energy, enabling a multi-channel imaging system where each discrete channel corresponds to a specific molecule type. Traditional systems have 4-5 channels/image; modern techniques allow up to 40-50 channels/image.

Once such images are acquired, a major problem comes in the task of interpreting them: how can one quantify the numbers and characteristics of different cells in an experiment? Here, I focus on the tasks of 1) identifying/segmenting individual cells within an image, and 2) creating cell-level image descriptors that capture heterogeneity between experimental conditions.

2.1: Review of existing work

Cell microscopy image interpretation is a well-studied field, with literature existing for > 50 years. Current approaches often use some features that are some combination of thresholding, gradients, or extracted features. From these features, a variety of segmentation algorithms have been applied, including graph cut, watershed, and level set optimization (figure 2).

A widely-held consensus opinion in the field is that the complexity of the segmentation algorithms required depends on the complexity of the image instances. Cells that do not overlap (Figure 2A) can be well segmented with thresholding alone, but cells that have complex overlap patterns (Figure 2C) require more complex approaches. Similarly, non-uniform cell brightness distributions (e.g. Figure 2D) require yet more sophisticated methods that can explicity model variability between cells as part of the segmentation.

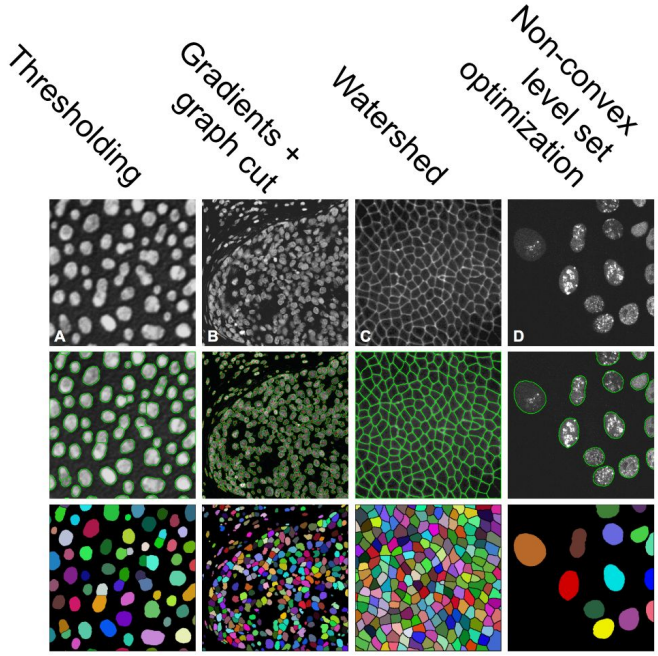


Figure 2: A survey of existing methods applied to progressively more complex cell segmentation tasks (Meijering et al., 2012).

2.2: Comparison to existing work

Multiscale Harris keypoint detectors have not been widely applied for the task of confocal microscopy image cell segmentation. While existing approaches do use multiscale detectors, many such detectors directly select the intrinsic scale for each pixel using the primary detector metric. In multiscale Harris keypoint detectors, and alternating optimization criteria is used, where keypoints are selected based on their maximization of the muti-scale LoG metric as well as the cornerness (weighted difference between determinant and squared trace) of the LoG second moment matrix. Although this iterative procedure takes longer to compute and will produce strictly fewer keypoints than multiscale LoG (since keypoints must be local LoG maxima and local Harris maxima rather than just LoG maxima), it has been shown to be less susceptible to occlusion and noise than LoG alone (Mikolajczyk, 2004). Occlusions and overlapping cells are common in immunofluroesence images, so it is reasonable to expect that better tolerance to these types of noise would improve performance. Additionally, multiscale Harris keypoint detectors are sensitive to intensity differences on local rather than global scales, so they offer improvements over threshold-based methods when images contain cells with different intensity scales (e.g. Figure 2D).

3.1: Technical Summary

Four methods for cell segmentation (thresholding with binary erosion, LoG, Harris, and multiscale Harris) and two methods for cell-level featurization (HoG and SIFT) are presented. I implemented each of these methods segmentation methods in Python, using no existing image filter methods for the percentile and LoG filters, and building Harris detectors starting from existing vectorized gaussian filters for performance. I implemented HoG in python, and used an existing SIFT implementation for comparison. I also wrote extensive code for evaluating performance, including summarizing bounding box accuracy compared to manual annotation, and classification accuracy for the image-of-origin task. I also wrote functions for image processing, including gamma correction for intensity normalization.

3.2: Technical description

Percentile based cell detector (baseline)

A commonly used method for cell detection in immunofluroesence images is a percentile threshold filter. This filter is parameterized with a percentile threshold *p* such that all pixel intensity values in an image with percentile < *p* are set to 0, and pixel intensity values > *p* are set to 1. I implemented an iterative percentile filter that uses a prior given by poisson(λ=50), and iterates *p* over the range [0.8, 1.0] to find a threshold that maximizes the posterior probability. In multiple comparisons, this approach significantly outperformed simple percentile thresholding because it searches for an appropriate threshold for each image and does not require setting a global limit *p*.

$$p(C = x) = Poisson(x; \lambda) = \frac{\lambda^x \exp -\lambda}{x!}$$

$$P(I; p) = \mathbb{1}[I_{ij} > S(I)_{|I|\frac{p}{100}}]$$

The percentile *p* is optimized according to these formulas, until a value is found that maximizes the prior on the number of detected cells. S is the magnitude sorted list of unique values in the image, where index *i* corresponds to the i^{th} largest value.

LoG keypoint based cell detector

Several published methods for cell segmentation use a Laplacian of Gaussian (LoG) filter to identify cells, which is more robust to occluded/overlapping cells than methods relying on gradients that surround each cell (e.g. methods using segmentation on global gradient alone or thresholding based methods). The LoG keypoint detector is a convolutional filter with weights parameterized by the second derivative of a gaussian distribution with scale σ_D.

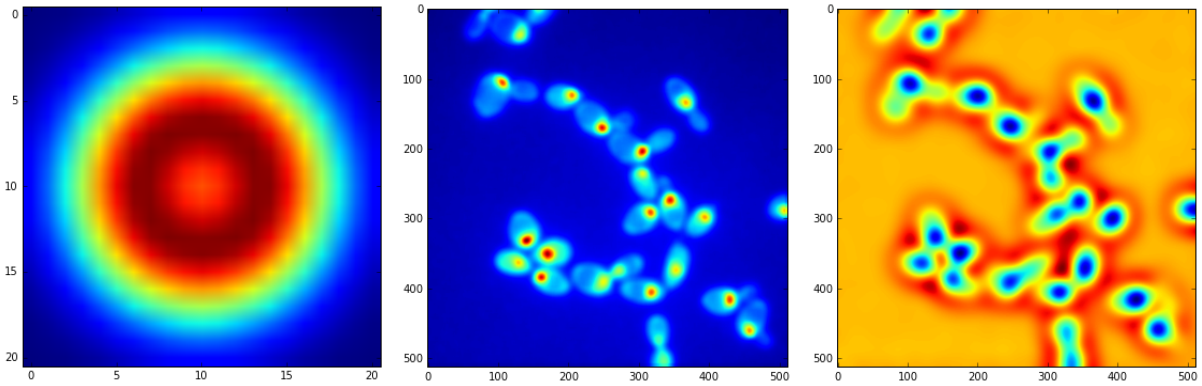


Figure: A LoG filter applied to a cell image. Left: the weights of a LoG with σ_D=7.5, center: the input image after pre-processing, right: the LoG activations, which show strong negative signal in the center of cells and strong positive signal immediately outside cells.

$$G(x, y, \sigma_D) = \frac{1}{2\pi\sigma_D^2} \exp \frac{-x^2 + y^2}{2\sigma_D^2}$$

$$L(x, y, \sigma_D) = I * G(x, y, \sigma_D)$$

$$LoG(x, y, \sigma_D) = \nabla L(x, y, \sigma_D) = L_{xx}(x, y, \sigma_D) + L_{yy}(x, y, \sigma_D)$$

$$LoG(x, y, \sigma_D) = \frac{1}{\pi\sigma_D^4}\left[1 - \frac{x^2 + y^2}{2\sigma_D^2}\right] \exp \frac{-x^2 + y^2}{2\sigma_D^2}$$

For notational consistency with the Harris methods, here the dispersion parameter of the gaussian distribution underlying the laplace is given as σ_D, which is the dispersion parameter in the Harris metric.

Multiscale LoG (mLoG) keypoint based cell detector

An extension of the Laplacian of Gaussian (LoG) filter is to use multiple scales, and then select the scale for a given pixel that corresponds to the highest LoG value. Since LoG filters are maximally activated by changes in gradient in multiple directions across a window equal to the filter dispersion parameter σ, objects with different characteristic gradient shifts produce maximal LoG scores under filters with different dispersion parameters. In the context of cell segmentation, this corresponds to cells of different sizes -- which do occur in the dataset, especially under different cell growth conditions.

$$mLoG(x, y, \sigma_D) = \sigma_D^2 \nabla L(x, y, \sigma_D) = \sigma_D^2 [L_{xx}(x, y, \sigma_D) + L_{yy}(x, y, \sigma_D)]$$

$$mLoG(x, y, \sigma_D) = \frac{1}{\pi\sigma_D^2}\left[1 - \frac{x^2 + y^2}{2\sigma_D^2}\right] \exp \frac{-x^2 + y^2}{2\sigma_D^2}$$

Harris detector, and multi-scale Harris detector

An extension of the multi-scale LoG detector is the Harris keypoint/corner detector, which also operates on the principle that at key points/corners, the image intensity value is expected to change significantly in multiple directions. Unlike in LoG, where we just score the sum of the second degree numerical gradient in the x and y directions, here we explicitly model the correlation and covariance between the gradients by also calculating the joint partial second derivative L_{xy}.

This higher order structure in the directional gradients is consistent with our model of cell physics: for a derivative smoothed over a distribution with standard deviation approximately equal to the diameter of the cell, we expect the gradient to change sharply in all directions, corresponding to the affine-transformed circular shape of cells.

The Harris metric uses as its input the structure tensor (or second moment matrix) of an image point, which describes the concordance in smoothed gradient direction from a given point. The tensor can be thought of as a gradient covariance matrix between image dimensions: for a 2-D image, the matrix is 2x2. The diagonal entries correspond to the second order numerical gradient in that dimension, and the off-diagonal entries correspond to the joint partial derivative between two dimensions. For 2-D images, this is the second order partial derivative in each of the x and y dimensions, and the joint partial derivative in the xy dimension. The gradients used to calculate the metric for a given point p_i are smoothed with a gaussian filter over a local window according to the dispersion parameter σ_D. The calculated Harris metrics for each pixel are then smoothed over a dispersion parameter σ_I. This joint smoothing adds an additional model parameter that must be optimized compared to LoG filtering, but reflects the signal property that gradients of individual cell edges and gradients of cell sizes are on different scales: cell membranes are thin and give abrupt signal changes but cells are large and have more diffuse gradient changes throughout. Therefore, I hypothesized that a multi-scale Harris detector may be able to better detect cells than a multi-scale LoG detector.

$$H(x, y, \sigma_D) = G(x, y, \sigma_D) \begin{bmatrix} L_{xx}(x, y, \sigma_D) & L_{xy}(x, y, \sigma_D) \\ L_{xx}(x, y, \sigma_D) & L_{yy}(x, y, \sigma_D) \end{bmatrix}$$

$$mH(x, y, \sigma_D, \sigma_I) = \sigma_D^2 G(x, y, \sigma_I) * \begin{bmatrix} L_{xx}(x, y, \sigma_D) & L_{xy}(x, y, \sigma_D) \\ L_{xx}(x, y, \sigma_D) & L_{yy}(x, y, \sigma_D) \end{bmatrix}$$

Multiscale HoG descriptor

I implemented a multiscale histogram of oriented gradients (HoG) descriptor to encode properties of an individual detected cell. The multiscale HoG method featurizes an image by calculating the intensity and angle of numerical gradients over a given distance. Pixel gradient directions are binned across a histogram with equal bin sizes, and weighted according to gradient magnitude. Descriptors are L-2 normalized, and written out for each block. The procedure is repeated for blocks of multiple sizes.

$$G_x(x,y) = I[x-1:x+1;y] \cdot [-1,0,1]^T$$

$$G_y(x,y) = I[x;y-1:y+1]^T \cdot [-1,0,1]^T$$

$$HoG(c_k = \{(x,y)_1,...,(x,y)_n\}) = H_{i\in 1...n}(arctan\left(\frac{G_y(x,y)_i}{G_x(x,y)_i}\right), G_y(x,y)_i + G_x(x,y)_i)$$

In HoG, G_x and G_y are the 1-point numerical gradients in the x and y directions respectively. They are equivalent to convolutional sobel filters in these directions. For simplicity, this formulation uses a weighted histogram function *H(x,w)*, which takes a value x to determine the bin, and a weight w to normalize. Additionally, I used L-2 normalization on each cell, which is not shown in the above formulation for simplicity.

SIFT descriptor

I used an existing implementation of the SIFT descriptor as a means to compare my implementation of the multiscale HoG descriptor for cell description. SIFT was used as a baseline, since it is currently a widely used method for generating cell descriptors for classification and analysis of confocal microscopy images.

PCA descriptor

As another benchmark featurization for experimental condition prediction, I used PCA featurization, which is the projection of each L-2 normalized image onto the top 10 principal components from PCA on the same L-2 normalized single cell image dataset.

4: Experiments

4: Experiments — present your experimental results of the method you have implemented with plots, graphs, images and visualizations. Show numerical/quantitative and qualitative results.

A set of 20,000 multi-channel immunofluoresence images were downloaded from the University of Washington's Yeast Resource Center (YRC; http://images.yeastrc.org/imagerepo/downloadImageRepo.do). Of these, I manually annotated cell bounding boxes for 10 randomly selected images (~110 cells). Each image in the dataset corresponds to a collection of yeast cells subject to the same experimental condition. The metadata provided with the database give parameters of the experimental conditions; the 10 hand-labeled images all correspond to different experimental conditions.

Task 1: Cell segmentation

Cell segmentation is the task of providing a bounding box for each cell in an image. It is possible to evaluate performance in different ways (detection rate, bounding box overlap proportion, centroid distance) depending on the target application. For the purpose of this project, I used bounding boxes, which are defined by four coordinate points and are the ordered connection of those four points.

I evaluated segmentation in 3 ways: qualitatively (by comparing properties of cell identification between methods across a variety of images), by proportion of cells identified (the proportion of

manually annotated cells that had a best-reciprocal-match annotation), and by the proportion of overlap (the proportion of pixels in the manual annotation that were also in the automated annotation). These reflect different characteristics of each algorithm's performance, and are also representative of real biologically motivated use cases. The ground truth in each comparison are hand-labeled cells.

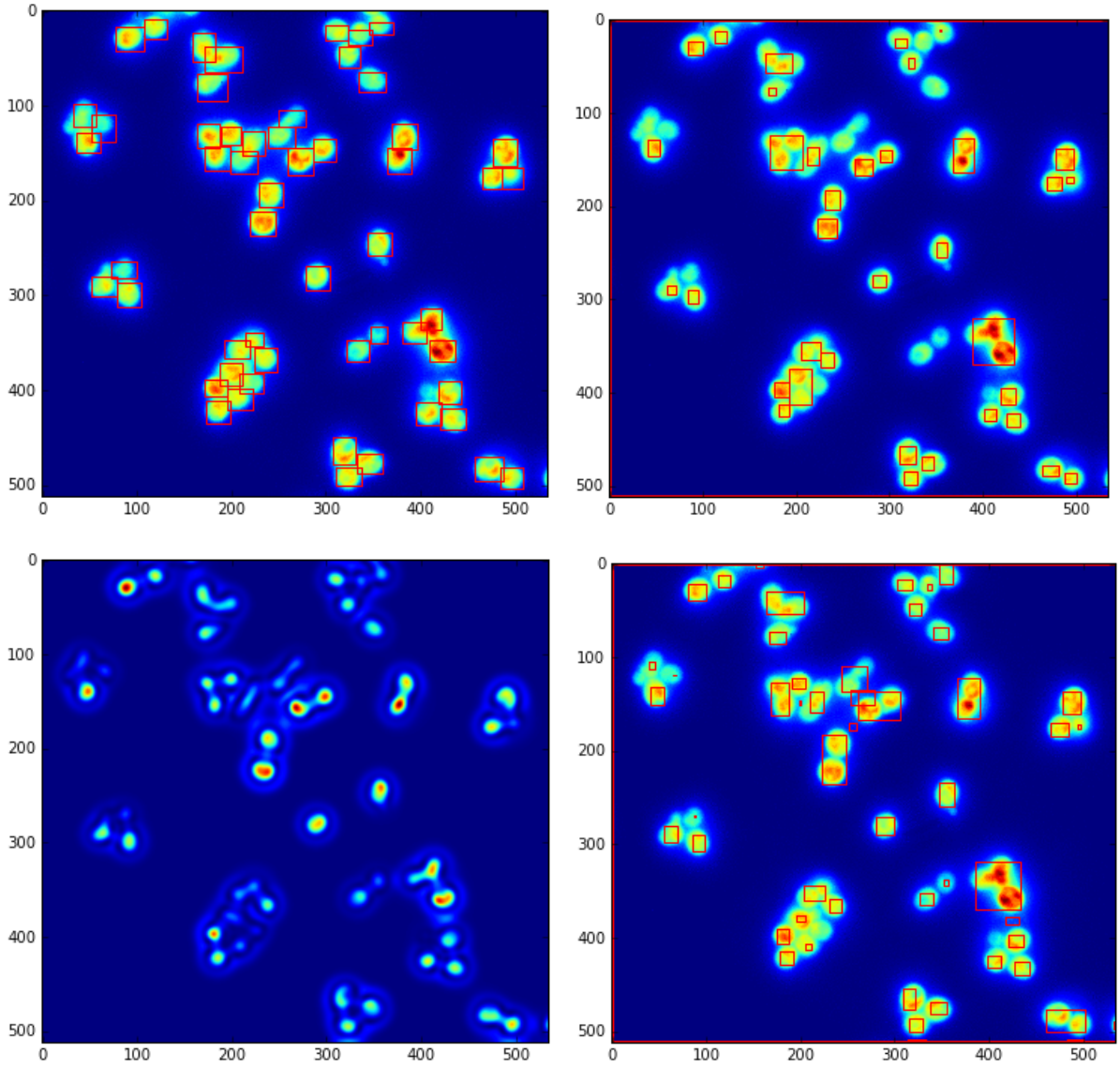
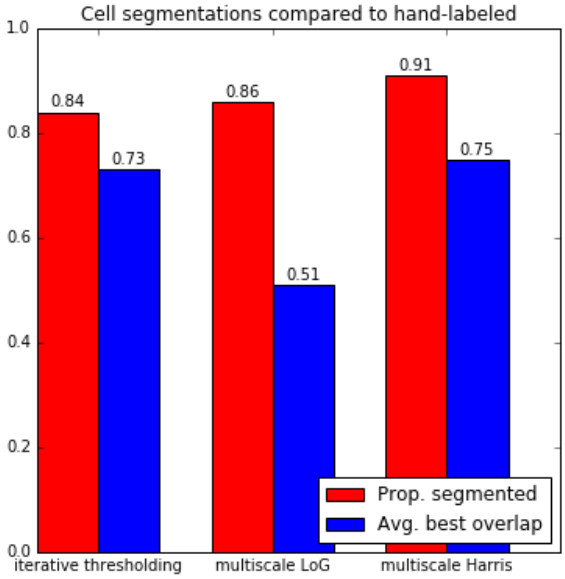


Figure: Cell segmentation example. Top left: processed image with manually annotated bounding boxes. Top right: result of iterative thresholding algorithm. Bottom left: Harris Affine feature map. Bottom right: a segmentation using Harris scores.

Overall, the Harris segmentation method outperformed the iterative thresholding and fixed thresholding baselines on the rate of bounding boxes detected. Part of the difference can be attributed to the fact that the Harris metric does a better job of identifying individual cells: for example, in the figures above, the cell cluster at approximately x=200, y=400 is correctly

segmented by the Harris detector (6/7 cells are annotated), but the thresholding method gets only 5/7 cells. Overall, either of these methods still outperform fixed thresholding, since setting the threshold high enough to segment cells within this cluster will prevent the filter from identifying cells with lower intensity elsewhere in the image.

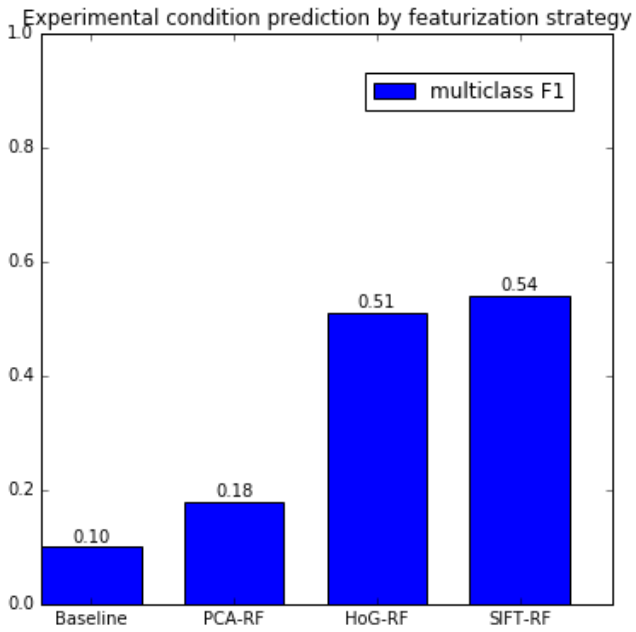


Task 2: Cell classification

Cell classification is the task of discriminating cell properties -- size, intensity distribution, shape, nuclear location, etc. Many instantiations of this task exist, however in most formulations it is important to recognize cells that are similar to each other from cells that are less similar. The dataset has the property that cells in the same image are exposed to the same experimental conditions, so cells from the same image have much higher similarity to each other than they have to cells from different images. On this basis, I use the task of image-of-origin prediction from normalized bounding boxes of cells.

Concretely, given original images $I_1,...,I_N$, which contain given cell bounding boxes C_{ij}, where index i corresponds to the image I_i and index j corresponds to the j^{th} cell bounding box in that image, the task is to predict the image label i corresponding to an unlabeled, L-2 normalized image subsection C_{ij}. For the purpose of this experiment, 10 images with at least 20 cells were selected, and a 0.9/0.1 training/evaluation split on a per-image basis was used. Fixed size 25x25 pixel cell bounding boxes were generated by thresholding with binary erosion, and positioning was tuned by hand.

Three different featurizations were used for classification: PCA, HoG, and SIFT. I used the version of L-2 normalized HoG that I implemented, and existing implementations of PCA and SIFT as comparisons to HoG. Weighted multi-class F1 on a 10-way prediction problem was used to evaluate performance. Overall, HoG and SIFT significantly outperformed PCA and baseline multiclass F1, and SIFT and HoG multi-class accuracies were directly comparable (0.51 and 0.54).



5: Conclusions

Keypoint-based cell segmentation can outperform threshold-based methods

Many existing algorithms for cell segmentation, e.g. the default algorithm used in the CellProfiler application, are threshold based. However, across a variety of image types (occlusion, cell vs. background intensity, etc.), I found that Harris detector based keypoint segmentation methods performed as well as or better than threshold based segmentations. Moreover, my implementation of the Harris detector was not significantly slower than my implementation of the iterative threshold based detector.

Multi-scale algorithms have better detection sensitivity and specificity than single-scale analogs

By experimenting with both single scale and multi-scale implementations of the LoG and Harris segmentation algorithms, I found that the multi-scale versions were able to identify more cells, and that the scale information found (e.g. the characteristic dispersion parameter for each local maxima) were informative about the correct size of the bounding box for that cell.

Featurization methods have significant impact on description accuracy

The HoG and SIFT featurizations significantly outperformed PCA and baseline on the image-of-origin task. This suggests that HoG and SIFT capture more of the intrinsic variability between images and better represent individual cells within an image. As an extension, it would be interested to compare embeddings of a deep convolutional autoencoder trained on individual cell images to the performance of these description methods.

Significant pre-processing is required before any segmentation or descriptor method works well

The images used in these experiments were all from the same database, where cells are imaged using the same experimental setup with the same microscope/camera. Still, a significant

amount of time was spent denoising images, throwing away brightfield (rather than fluorescent) images, and experimenting with image adjustments such as gamma correction. Had the images not all come from the same dataset (which would make the methods more adaptable), pre-processing would have taken much longer: I was largely able to reuse processing parameters between images, but likely wouldn't have been able to if images had significantly different intensity distributions.

6: References

[1] "Cell Segmentation: 50 Years Down the Road." [Online]. Available: http://www.imagescience.org/meijering/publications/1095/. [Accessed: 04-Jun-2016].

[2] X. Lou, U. Koethe, J. Wittbrodt, and F. A. Hamprecht, "Learning to segment dense cell nuclei with shape prior," in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1012–1018.

[3] Z. Yin, R. Bise, M. Chen, and T. Kanade, "Cell segmentation in microscopy imagery using a bag of local Bayesian classifiers," in *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2010, pp. 125–128.

[4] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. Kang, O. Friman, D. A. Guertin, J. Chang, R. A. Lindquist, J. Moffat, P. Golland, and D. M. Sabatini, "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," *Genome Biology*, vol. 7, p. R100, 2006.

[5] P. Yan, X. Zhou, M. Shah, and S. T. C. Wong, "Automatic Segmentation of High-Throughput RNAi Fluorescent Cellular Images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 109–117, Jan. 2008.

[6] O. Dzyubachyk, W. Niessen, and E. Meijering, "Advanced level-set based multiple-cell segmentation and tracking in time-lapse fluorescence microscopy images," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2008, pp. 185–188.

[7] X. Chen, X. Zhou, and S. T. C. Wong, "Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 4, pp. 762–766, Apr. 2006.

[8] A. Kuijper and B. Heise, "An automatic cell segmentation method for differential interference contrast microscopy," in *19th International Conference on Pattern Recognition, 2008. ICPR 2008*, 2008, pp. 1–4.

[9] O. Z. Kraus, L. J. Ba, and B. Frey, "Classifying and Segmenting Microscopy Images Using Convolutional Multiple Instance Learning," *arXiv:1511.05286 [cs, q-bio, stat]*, Nov. 2015.