CHARACTERIZING MORPHOLOGY DIFFERENCES FROM IMAGE DATA USING A MODIFIED FISHER CRITERION

Wei Wang[1]*, Yilin Mo*[2]*, John A. Ozolek*[3]*, Gustavo K. Rohde*[1,2,4]

[1]Center for Bioimage Informatics, Department of Biomedical Engineering
[2] Department of Electrical and Computer Engineering
[4] Lane Center for Computational Biology
Carnegie Mellon University, Pittsburgh, PA. 15213
[3] Department of Pathology, Children's Hospital of Pittsburgh, Pittsburgh, PA. 15201

ABSTRACT

Image-based morphometry of cells, tissues, and organs is an important topic in biomedical image analysis. We propose a novel method to characterize the morphological information that discriminates between two populations of morphological exemplars (cells, organs). We first demonstrate that the application of standard techniques such as Fisher linear discriminant analysis (FDA) can lead to undesirable errors in characterizing such information. We then describe an adaptation of the FDA technique that utilizes a least squares projection error to regularize the final solution. We show results comparing the FDA, modified FDA, and principal component analysis (PCA) techniques utilizing a contour-based characterization of both simulated and real images of cell nuclei.

Index Terms— Fisher linear discriminant analysis, Morphological analysis, Data visualization

1. INTRODUCTION

Clinicians, biologists, and other researchers have long used information about shape, form, and texture to understand biological differences between different biological structures [1, 2, 3]. Image-based quantitative morphology is concerned with the application of statistical analysis techniques to extract important information that characterizes the distribution of such data. Early work often focused on numerical feature-based approaches (e.g. measuring size, form factor, etc.) that aim to quantify and measure differences between different forms in carefully constructed feature spaces [3, 4]. Modern approaches to this problem have shifted to viewing the entire morphological exemplar (as depicted in a biomedical image) as a point in a carefully constructed metric space [5, 6, 7]. Advantages of such approaches include the fact that little or no reduction of information is involved (the entire image information can often be utilized to compute the metric) and that deformations mapping one form to another can often be computed, thus facilitating the application of geometric techniques to understand the distribution of a given dataset.

At the microscopic level (cells and subcellular components) contours and medial axis representations for each structure can often be extracted and the PCA related techniques employed for characterizing the distribution of each population of exemplars (e.g. type of cell) [8, 9, 10]. Our group has also experimented with the application of deformation and transportation-based metrics, together with techniques such as multidimensional scaling and FDA, to characterize distributions of microscopic images of cell nuclei [11, 12]. At the macroscopic level several groups have utilized medial axis and deformation-based metrics to quantify the differences between two anatomical exemplars [7, 13, 14]. In addition, when a linear embedding of the data is assumed the PCA technique is often utilized to understand the main modes of variation present in the data [13, 15, 16].

While useful for visualizing main modes of variation for a group of cells, sub-cellular structures, or organs, the PCA technique is often inconclusive when applied to quantify differences in form between two or more populations. Linear discriminant analysis techniques (e.g. FDA [3]), on the other hand, are specifically designed for computing a linear subspace representation (arranged in order of importance) that directly measures how groups of structures differ according to a given metric. In this work, however, we show that when applied to real image data, such discriminant techniques can lead to erroneous interpretation about the differences in form that are actually present in the data. We propose a modified Fisher criterion by combining the traditional Fisher discrimination metric with a least squares projection error and show that this leads to a regularized eigen decomposition problem. We show the application of the technique in contours obtained from real and simulated datasets and demonstrate how our technique can help reduce some of the errors encountered when applying the standard FDA technique.

2. METHODS

2.1. Linear metric space-based analysis of biomedical images

The method we describe can be applied whenever a linear embedding for the image data can be assumed and obtained. That is, given an image I_i depicting a cell or organ, a one-to-one mathematical function f can be computed to map each image to a linear space. Mathematically: $f(I_i) = \mathbf{x}_i$, with $\mathbf{x}_i \in \mathcal{R}^m$, with m the dimension of the linear space. In addition, we assume the function g that maps any coordinate $\mathbf{x}$ in the linear space back to the unique corresponding image (biological form) is also known. That is $I = g(\mathbf{x})$. In this work we utilize contours obtained from image data as the linear representation of each image. However, other linear embeddings could also be utilized [13, 15, 16].

In the computations we present to utilize the contour-based metric [9, 10, 17] to characterize the shape of nuclear image data. Briefly, after segmentation and pre-processing (see section 3.2 for details), a binary mask containing each nucleus is obtained. The mask is then eroded by one pixel and the result subtracted from the initial mask to obtain pixels depicting the contour of each nucleus. The contour is then converted to a polar coordinate system with respect to the center of the contour, and n points are sampled (separated by equally

distant angles) to obtain a parametric representation of each image. This procedure maps each image I_i to a point $\mathbf{x}_i$ in the standard $\mathcal{R}^{2n}$ vector space. As our purpose here is to quantify shape, a point $\mathbf{x} \in \mathcal{R}^{2n}$ can be plot on the same coordinate system as the given image data to obtain a visualization of the corresponding shape.

2.2. Fisher discriminant analysis

Given a set of data points $\mathbf{x}_i$, for $i = 1, \cdots, N$, with each index i belonging to class c, the problem proposed by Fisher [3, 18] relates to solving the following optimization problem

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} \tag{1}$$

where $S_B = \sum_c N_c(\boldsymbol{\mu}_c - \bar{\mathbf{x}})(\boldsymbol{\mu}_c - \bar{\mathbf{x}})^T$ represents the "between class scatter matrix", $S_W = \sum_c \sum_{i \in c}(\mathbf{x}_i - \boldsymbol{\mu}_c)(\mathbf{x}_i - \boldsymbol{\mu}_c)^T$ represents the "within classes scatter matrix", $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$ represents center of the entire data set, N_c is the number of data in class c and $\boldsymbol{\mu}_c$ is the center of class c. The maximization problem above is equivalent to the following problem [18]:

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} J(\mathbf{w}) = \frac{\mathbf{w}^T S_T \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} \tag{2}$$

where $S_T = \sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$ represents the "total scatter matrix", and $S_T = S_B + S_W$. The criterion is then optimized by solving the generalized eigenvalue problem [18] $S_T \mathbf{w} = \lambda S_W \mathbf{w}$, and selecting the eigenvector associated with the largest eigenvalue. In our paper, we focus on the two-class problem, the above discussion is limited to computing a single "most discriminating" direction. In a multiple-class problem, a basis for a linear discriminating subspace is sought, the criterion (2) can be modified and the solution is given by the "top" eigenvectors of the generalized eigenvalue problem. Usually, we subtract each data by this mean $\mathbf{x}'_i = \mathbf{x}_i - \bar{\mathbf{x}}$ before we compute the scatter matrices S_W, S_B, S_T.

2.3. A simulated data example

We simulated two classes of circle shapes (100 circles for each class) with different radii. For one class, the radii were uniformly distributed from 0.42 to 0.62 pixels while for the other class, radii ranged from 0.28 to 0.48 (also uniformly distributed). If we use n sample points on the contour, each circle can be mapped to a point in space R^{2n}. The data, however, occupies a linear one dimensional subspace of R^{2n}, since the points depicting any circle can be computed by a linear combination between any two circles with different radii. Because only one parameter (radius) varied in our simulation, this linear subspace is one dimensional. For the purpose of visualizing the concepts we are about to describe, each data point (circle) can then be uniquely mapped to the two dimensional vector space shown in Figure 1 where the $Y1$ coordinate represents the y coordinate of the top most sample point on the circle, while the $Y2$ coordinate represents the y coordinate bottom most sample point. From the coordinates $\mathbf{x}_i = (Y1_i, Y2_i)^T$ any circle can be reconstructed.

The solution $\mathbf{w}^*$ of the FDA problem discussed above can be visualized by plotting $\mathbf{x}_\gamma = \bar{\mathbf{x}} + \gamma \mathbf{w}^*$ for some range of b. Figure 1(A) contains the circles correspinding to $\mathbf{x}_\gamma$ for $-3\sigma \leq \gamma \leq 3\sigma$, where σ is the standard deviation (square root of eigenvalue ι). Visual inspection of the results in Figure 1(A) quickly reveals the problem. The method indicates that circle *translation* in combination of a change in size is the geometric variation that best separates the

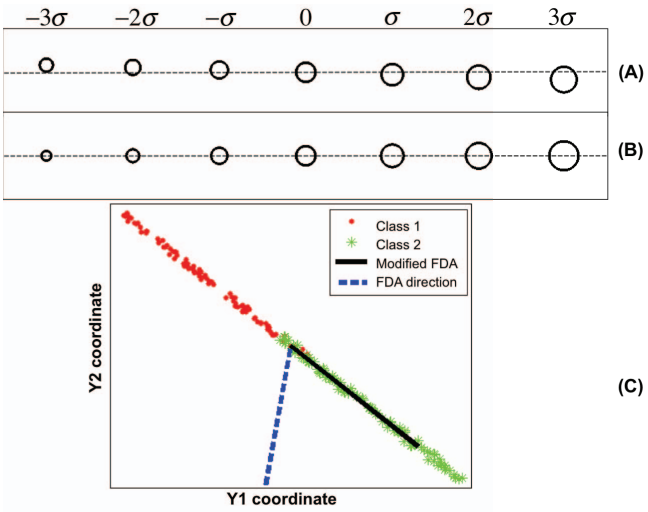

Fig. 1. Discriminant information computed for simulated data set. For (A) and (B), each column shows the result of adding mean shapes with different variations along the computed directions. A: Visualization of computed most discriminant direction by directly applying FDA. B: Visualization of computed most discriminant direction by our modified FDA method. C: Plot of two sample points on the contour for the whole data set (Please see text for detail).

two distributions according the Fisher criterion. While such a direction may allow for high classification accuracy, by construction, the data contained no variation in the position (translation) of the circles. We can therefore understand that such results are misleading, since the translation effect is manufactured by the FDA procedure and does not exist in the data. The problem is illustrated in part C of the figure, where the two distributions are plotted, together with the solution by the Fisher method (dotted blue line): the distribution of datapoints do not lie on or near the direction computed by FDA.

2.4. A modified FDA criterion

The FDA criterion can be modified by adding a term that "penalizes" directions $\mathbf{w}$ that do not pass close to the data. More specifically, we propose to combine the standard FDA criterion with a term that measures the distance of each data point to the computed direction $\mathbf{w}$. Mathematically, an arbitrary line in the shape space R^{2n} can be represented as $\lambda\mathbf{w} + \mathbf{b}$, with line direction and offset $\mathbf{w}, \mathbf{b} \in R^{2n}$, $\lambda \in R$. The distance from a data point $\mathbf{x}_i$ in the shape space R^{2n} to the line can be found by solving λ in the following optimization problem, $\min_\lambda \|\lambda\mathbf{w} + \mathbf{b} - \mathbf{x}_i\|^2_{R^{2n}}$. The solution is $\lambda_i^* = -(\mathbf{b} - \mathbf{x}_i)^T\mathbf{w}/(\mathbf{w}^T\mathbf{w})$. Therefore, the squared distance d_i from the point $\mathbf{x}_i$ to the line is $d_i^2 = \|\lambda_i^*\mathbf{w} + \mathbf{b} - \mathbf{x}_i\|^2$. Substituting λ_i^*, we have:

$$d_i^2 = (\mathbf{b} - \mathbf{x}_i)^T(\mathbf{b} - \mathbf{x}_i) - \frac{(\mathbf{b} - \mathbf{x}_i)^T\mathbf{w}\mathbf{w}^T(\mathbf{b} - \mathbf{x}_i)}{\mathbf{w}^T\mathbf{w}}$$

$$= tr\left[(\mathbf{b} - \mathbf{x}_i)(\mathbf{b} - \mathbf{x}_i)^T\left(\mathbf{I} - \frac{\mathbf{w}\mathbf{w}^T}{\mathbf{w}^T\mathbf{w}}\right)\right].$$

For a data set of N points, the sum of squared distances from each point in that data set to that line is:

$$\sum_{i=1}^{N} d_i^2 = \sum_{i=1}^{N} tr\left[(\mathbf{b}-\mathbf{x}_i)(\mathbf{b}-\mathbf{x}_i)^T\left(\mathbf{I}-\frac{\mathbf{w}\mathbf{w}^T}{\mathbf{w}^T\mathbf{w}}\right)\right] \quad (3)$$

Recall that our goal is to maximize the Fisher criterion defined in equation (2) while minimizing the sum of squared distances defined in equation (3) to guarantee the line is well populated by the data. We note that the Fisher criterion is independent of the offset $\mathbf{b}$ and the term defined in equation (3) contains $\mathbf{b}$ multiplying the terms containing $\mathbf{w}$. Since it should be minimum for all possible choices of $\mathbf{w}$, $\mathbf{b}$ can be chosen independently of $\mathbf{w}$ and can be shown to be (we omit details of the derivation for brevity): $\mathbf{b}^* = \sum_{i=1}^{N} \mathbf{x}_i/N$. This indicates that this line must go through the center of the data set. Equation (3) can then be written as:

$$\min_{\mathbf{w},\mathbf{b}=\mathbf{b}^*} \sum_{i=1}^{N} d_i^2 = \min_{\mathbf{w},\mathbf{b}=\mathbf{b}^*} \left\{ tr(S_T) - \left(\frac{\mathbf{w}^T S_T \mathbf{w}}{\mathbf{w}^T\mathbf{w}}\right)\right\} \quad (4)$$

where $S_T = \sum_{i=1}^{N} tr\left[(\mathbf{b}^*-\mathbf{x}_i)(\mathbf{b}^*-\mathbf{x}_i)^T\right]$ is the same "scatter matrix" as in equation (2). The optimization problem defined in equation (4) is equivalent to:

$$\min_{\mathbf{w}} \left\{ -\left(\frac{\mathbf{w}^T S_T \mathbf{w}}{\mathbf{w}^T\mathbf{w}}\right)\right\} \quad (5)$$

Both equation (4) and equation (5) will give a big value if the line go far away from the data populations. To simplify our solution (see section 2.5), we use equation (5) as the penalty term.

We know that maximizing the Fisher criterion defined in equation (2) is equivalent to maximizing $-\frac{1}{J(\mathbf{w})}$ [3, 18]. To maximize the fisher criterion (discriminating information) and minimize the penalty term defined in equation (5) at the same time, we can define the modified Fisher criterion as follows:

$$\max_{\mathbf{w},\mathbf{b}} E(\mathbf{w},\mathbf{b}) = \max_{\mathbf{w},\mathbf{b}} \left\{ -\frac{1}{J(\mathbf{w})} + \alpha * penalty \right\} \quad (6)$$

$$max_{\mathbf{w}} \left\{ -\frac{\mathbf{w}^T S_W \mathbf{w}}{\mathbf{w}^T S_T \mathbf{w}} - \alpha \frac{\mathbf{w}^T\mathbf{w}}{\mathbf{w}^T S_T \mathbf{w}} \right\} \quad (7)$$

where α is a scalar weight term.

2.5. Solution for modified FDA

The optimization problem defined in equation (7) can be simplified as the following optimization problem:

$$max_{\mathbf{w}} \left\{ \frac{\mathbf{w}^T S_T \mathbf{w}}{\mathbf{w}^T \left(S_W + \alpha \mathbf{I}\right) \mathbf{w}} \right\} \quad (8)$$

where $\mathbf{I}$ is the identity matrix. The solution for the problem above is also given by a generalized eigenvalue decomposition $S_T\mathbf{w} = \lambda\left(S_W + \alpha\mathbf{I}\right)\mathbf{w}$. This solution is similar to the solution of the traditional FDA problem, with the regularization provided by $\alpha\mathbf{I}$ [18, 19]. However, we provide a geometric interpretation of this regularization term, which is, as described above, the minimization of the least squares projection error.

3. RESULTS

3.1. Simulated experiment

We test the modified FDA method above on the simulated dataset depicted in Figure 1. We can compare the result of applying the FDA method (Fig 1 part A) with our modified FDA method (with $\alpha = 0.01$) in Fig 1(B). We can see the method we propose does indeed recover the correct information that discriminates between the two populations (in this case, the radii of the circles). While this is not necessarily the most discriminating information, it is the most discriminating information that is well populated by the data, in the sense made explicit by equation (6). In this specific simulation the modified FDA method yields the same result as the standard PCA method would. However, as shown in the real data examples below, that is not a general rule.

3.2. Real data experiments

We also apply our method on a real biomedical image data set to quantify the difference, in nuclear morphology, between normal versus cancerous nuclei. The raw data consisted of histopathology images originating from five cases of liver hepatoblastoma (HB), each containing adjacent normal tissue (NL). The data was taken for the archives at the Children's Hospital of Pittsburgh, and is described in more detail in [12, 20]. The images were segmented by a semi automatic method involving a level set contour extraction. They were normalized for translation, rotations, and coordinate inversions as described in our earlier work [11, 12, 20]. The dataset we use here consisted of 500 nuclear contours: 250 for (NL), and 250 for (HB). Each image I_i was mapped to a 180 dimensional vector $\mathbf{x}_i$.

In Fig 2(A), we demonstrate the first two modes of variations computed by PCA. The first modes of variation is size and the second is the elongation of the nuclei. In Fig 2(B), we demonstrate the discriminating mode computed by our modified FDA (with $\alpha = 0.01$). For this specific cancer, we can see that size of the nuclei as well as the protrusion and invagination of the nuclei are the discriminating information. As in section 3.1, we also directly apply FDA on the contours of data set. Some sample points on contours start to move perpendicularly to the contour with relatively big variation, while some remain unchanged. The direction computed by FDA does not seem to capture visually interpretable information.

4. SUMMARY AND DISCUSSION

Quantifying the information that is different between two groups of cells or organs is an important problem in biology and medicine. We have shown that the application of the standard FDA criterion to solving this problem (other linear discrimination methods can also suffer from the same shortfalls) can lead to erroneous results in interpretation. To solve this problem, we proposed a novel method to characterize the morphological information that discriminates between two populations of morphological exemplars (cells, organs). The method is based on the solution of a FDA criterion by adding a penalty term composed of sum of squared distances from each point in that data set to that line. The set of solutions given by our method is dependent on the parameter α that weights the combination of the two terms. It can be shown that when $\alpha = 0$ the solution is given by the traditional FDA method, while as $\alpha \to \infty$ the solution approximates the solution of the standard PCA method. We also showed that the solution of the modified criterion is given by a "regularized" eigen decomposition problem. While others have also used

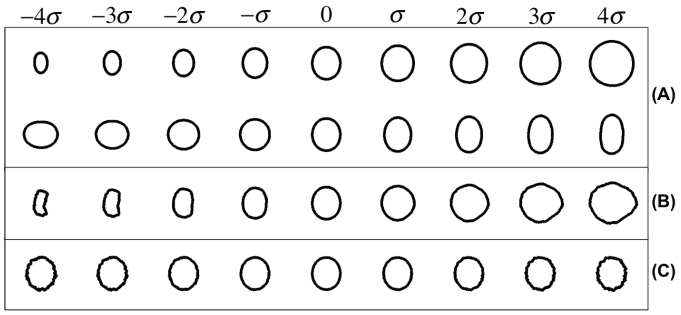

Fig. 2. Principle variations and discriminant information computed for real liver nuclei data. Each column shows the result of adding mean shapes with different variations along the computed directions. A: First 2 principle variations computed by Principle Component Analysis (PCA). B: Discriminant variation computed by our modified FDA method. We can see that size difference combined with the protrusion and invagination of the nuclei are the discriminating information. C: direction computed by the standard FDA procedure. See text for more details.

the same regularized solution (see [19] for examples) we show that this is equivalent to maximizing the FDA criterion while minimizing the error of the projection of the data. The regularization also adds the convenience that it facilitates the numerical solution of the associated generalized eigenvalue problem.

Results on simulated and real data confirm that our modified FDA helps overcome the limitations of the standard FDA method. In particular, the application of the standard FDA method to real nuclear data seems to yield solutions that are far from being closed contours. Results generated by applying our modified FDA criterion on real liver nuclei data indicate that, in this specific liver cancer, the size difference combined with the protrusion and invagination of the nuclei represents the most discriminating information between these two sets of nuclei, that is actually present in the data.

Finally, we emphasize that although we have used contours extracted from image data as our linear embeddings, it is possible to use the same method on other linear embeddings [13]. For some such linear embeddings, however, distance measurements, projections over directions, etc., over large distances (large deformations) may not be appropriate. In such cases we believe the same modified FDA method could be used locally, in an idea similar to that presented in [21].

5. ACKNOWLEDGEMENT

This work was partially supported by NIH grant 5R21GM088816.

6. REFERENCES

[1] G. Papanicolaou, "New cancer diagnosis," *CA: A Cancer Journal for Clinicians*, vol. 23, no. 3, p. 174, 1973.

[2] J. Thomson, "On growth and form," *Nature*, vol. 100, pp. 21–22, 1917.

[3] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, pp. 179–188, 1936.

[4] J. Prewitt and M. Mendelsohn, "The analysis of cell images," *Annals of the New York Academy of Sciences*, vol. 128, no. 3, pp. 1035–1053, 1965.

[5] D. G. Kendall, "Shape manifolds, procrustean metrics, and complex projective spaces," *Bull Lond Math Soc*, vol. 16, pp. 81–121, 1984.

[6] F. L. Bookstein, *The Measurement of Biological Shape and Shape Change*. Springer, 1978.

[7] U. Grenander and M. I. Miller, "Computational anatomy: an emerging discipline," *Quart. Appl. Math.*, vol. 56, no. 4, pp. 617–694, 1998.

[8] H. Blum *et al.*, "A transformation for extracting new descriptors of shape," *Models for the perception of speech and visual form*, vol. 19, no. 5, pp. 362–380, 1967.

[9] Z. Pincus and J. A. Theriot, "Comparison of quantitative methods for cell-shape analysis," *J Microsc*, vol. 227, no. Pt 2, pp. 140–56, Aug 2007.

[10] T. Zhao and R. F. Murphy, "Automated learning of generative models for subcellular location: building blocks for systems biology," *Cytometry A*, vol. 71A, pp. 978–990, 2007.

[11] G. K. Rohde, A. J. S. Ribeiro, K. N. Dahl, and R. F. Murphy, "Deformation-based nuclear morphometry: capturing nuclear shape variation in hela cells," *Cytometry A*, vol. 73, no. 4, pp. 341–50, Apr 2008.

[12] W. Wang, J. A. Ozolek, D. Slepcev, A. B. Lee, C. Chen, and G. K. Rohde, "An optimal transportation approach for nuclear structure-based pathology," *IEEE Trans Med Imaging*, 2010.

[13] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration," *IEEE Trans. Med. Imaging*, vol. 22, no. 8, pp. 1014–1025, 2003.

[14] P. T. Fletcher, C. L. Lu, S. A. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans. Med. Imag.*, vol. 23, pp. 995–1005, 2004.

[15] M. Vaillant, M. Miller, L. Younes, and A. Trouvé, "Statistics on diffeomorphisms via tangent space representations," *NeuroImage*, vol. 23, pp. S161–S169, 2004.

[16] S. Makrogiannis, R. Verma, and C. Davatzikos, "Anatomical equivalence class: A morphological analysis framework using a lossless shape descriptor," *IEEE Trans. Med. Imaging*, vol. 26, no. 4, pp. 619–631, 2007.

[17] T. Cootes, C. Taylor, D. Cooper, J. Graham *et al.*, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[18] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, August 2006.

[19] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Mullers, "Fisher discriminant analysis with kernels," in *Proceedings of the 1999 IEEE Signal Processing Society Workshop, Neural Networks for Signal Processing IX*, 2002, pp. 41–48.

[20] W. Wang, J. Ozolek, and G. Rohde, "Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images," *Cytometry Part A*, vol. 77, no. 5, pp. 485–494, 2010.

[21] H. Zhang, A. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 2126–2136, 2006.