Automatic Image Segmentation and Classification Using On-line Shape Learning

Kyoung-Mi Lee
Computer Science Department
The University of Iowa
Iowa City, IA 52242

W. Nick Street
Management Sciences Department
The University of Iowa
Iowa City, IA 52242

Abstract

The detection, precise segmentation and classification of specific objects is an important task in many computer vision and image analysis problems, particularly in medical domains. Existing methods such as template matching typically require excessive computation and user interaction, particularly if the desired objects have a variety of different shapes. This paper presents a new approach that uses unsupervised learning to find a set of templates specific to the objects being outlined by the user. The templates are formed by averaging the shapes that belong to a particular cluster, and are used to guide an intelligent search through the space of possible objects. This results in decreased time and increased accuracy for repetitive segmentation problems, as system performance improves with continued use. Further, the information gained through clustering and user feedback is used to classify the objects for problems in which shape is relevant to the classification. The effectiveness of the resulting system is demonstrated on two applications: a medical diagnosis task using cytological images and a vehicle recognition task.

1. Introduction

Quantitative analysis of digital images requires detection and segmentation of the borders of objects of interest. For instance, in medical images, this detection and segmentation play important roles in diagnosis and prognosis. Unfortunately, even with the aid of computer vision software, traditional manual analysis is tedious and time consuming, especially in cases where a large number of objects must be specified. While many commercial systems include new and sophisticated analysis capability to circumvent these problems [2, 15], they are often limited by their narrow focus and typically rely heavily on user corrections to specify the exact border, a necessary step for precise analysis. Thus the development of highly efficient and robust techniques to automatically and quickly segment an exact border is an important goal.

Many applications of image recognition require an object to be classified. Consider the problem of isolating cell nuclei from heterogeneous tissue for dissection and genetic analysis. Such tissue may contain both diseased and healthy cells in the same sample. However, the molecular analysis may depend on collecting a "clean" sample of all diseased or all healthy cells for comparison purposes. Further, the user should not be expected to wait for the system to locate objects that are not desired for the particular experiment.

There are several methods to segment and classify objects in images. Korn *et al.* used a natural similarity function for shape matching, based on concepts from mathematical morphology, and k-nearest-neighbor search for medical tumor shape classification [9]. Dubuisson *et al.* introduced an algorithm to extract the contour of moving vehicles and generic models for five types of vehicles represented by line segments [6]. Since the imaging device was at fixed position, image subtraction-based segmentation was used to estimate the initial location of a vehicle of interest.

These problems have been addressed with machine learning methods as a means for improving knowledge used in the imaging process, and thus for producing more robust software [5, 12]. A key idea in applied machine leaning involves building an advisory system that recommends actions but gives final control to a human expert, with each decision generating a training case, gathered in an unobtrusive way, for use in learning. This interactive approach promises to greatly reduce the number of decisions that the system must make per image, thus improving the ability to deal with a high flow of images. Moreover, the resulting systems should adapt their knowledge to the preferences of individuals in response to feedback from those users.

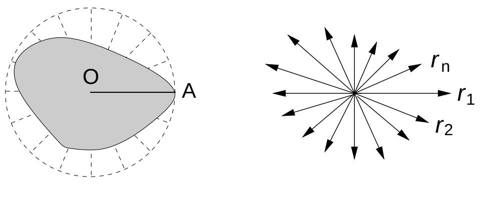
In this paper, we will focus on learning shapes to detect quickly and robustly and to segment accurately. While the shapes can be represented by various models, regardless of the shape model used, it is usually desirable for a model to describe an average shape to contain information about shape variation within the training set [7]. The proposed system uses an on-line learning method ideal for highly repetitive tasks such as morphological analysis of cytological and histological images, in which many cells or cell nuclei must be precisely outlined in many different images, and the shapes of the individual cells vary widely from one sample to the next. Our platform for medical image analysis is the Xcyt program [13, 16], a graphical computer program for diagnosing breast cancer and predicting the course of the disease based on nuclear morphometry. However, our approach is not limited to the medical domain. By using a very general shape model, our system can learn to segment and classify a wide variety of objects.

The remainder of the paper is organized as follows. Section 2 describes our shape model. Section 3 describes the proposed system, including the learning of templates. Experimental results in Section 4 show that the quality of segmentation produced using the proposed algorithm is comparable to an exhaustive template search, and the computation time is significantly reduced. Conclusions and future work are presented in Section 5.

2. Modeling Shapes

In order to model a shape, we adopt the centroid-radii model to represent the shape [3, 11]. Assume that OA is an arbitrary radius of the template. In our algorithm, starting from OA and moving clockwise, we divide the circle into n equal arcs to place points around the boundary, with the regular interval being $\frac{360}{n}$ degrees. So the shape can be represented as

$$\mathbf{v} = (r_1, r_2, \cdots, r_n).$$



(a) Object (b) Centroid-radii model

Figure 1: Modeling of shape

The shape shown in Figure 1 is represented with 16 linked nodes. Translation invariance is achieved through the use of an object-centered coordinate system. Shapes are normalized for size, orientation and reflection during the algorithm, as described in Section 3.3. This representation, motivated by its ease of use template matching using a generalized Hough transform (GHT) [1], allows the matching of a wide variety of topologically simple shapes.

3. Description of the system

Figure 2 shows a flowchart of the proposed system. The system consists of four parts: detecting, segmenting, clustering and classifying. While these four parts are performed repeatedly, the system learns and gathers knowledge of shapes the user wants to find.

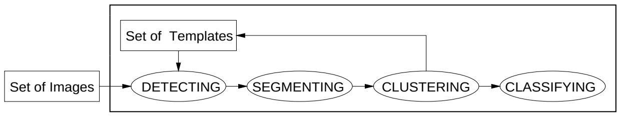


Figure 2: Flowchart of the automatic system

3.1. Initializing Templates

Initialization of the templates can proceed in one of two ways, depending on whether *a priori* knowledge about the shapes is available. In the first application, the objects in question are human cell nuclei, which are more or less elliptical in shape. Therefore we can initialize a set of elliptical templates with different shapes. Since the algorithm is invariant to size, orientation and reflection, only one template with each shape is created. The initial templates are created automatically but could easily be drawn by the user. In the second application, we assume no knowledge of the shapes, so the initial templates are created by the user outlining the first few objects with a mouse.

3.2. Detecting and Segmenting Objects

GHT [1] is a typical algorithm for detecting objects in images. Generally, an array of points, of the same size as the input image, comprises the 2D accumulator, in which each point has a value specifying the possibility that the reference point of an object shape to be detected is located at the point. These values are accumulated by having each edge point in the image "vote" for the positions that could correspond to the reference point of the particular template. To perform GHT, the template shapes are built in advance, requiring the user to know exactly what shapes will be encountered. Further, when the scale and orientation of an input shape are variant and unknown in advance, brute force is usually employed to enumerate all possible scales and rotations of the input shape in the GHT process, thus adding dimensions to the accumulator.

We use a GHT variant known as Iterative GHT (IGHT) [14] which eliminates the extra dimensions by using a local accumulator (to find each template in turn) and a global accumulator (to collect the best scores from the local accumulator). For each new image, IGHT is performed using the existing set of shape templates at various scale factors. After plateaus in the accumulator are removed with a peak-sharpening step, the highest value in the global accumulator is found. The appropriate template is overlaid onto the image at the appropriate rotation angle. This process is repeated until a user-defined number of objects has been located.

An important aspect of the proposed system is the use of intelligent search to guide the order in which template matches are attempted. As the images are processed, the program records the number of times each template has matched an object and searches first for the objects that we are most likely to find. This significantly reduces the expected time required, since not all templates will appear in every image.

Previous work [10, 16] has shown that the system isolates the cell nuclei very well using the combination of IGHT and snakes [8], which are initialized from the matching templates. The user can then edit or remove a boundary, or draw a boundary on an undetected object by hand using the mouse. From the perspective of the learning system, this process creates a collection of positive examples of the shapes that this user would like to find and outline.

3.3. Clustering Shapes

After an object has been correctly outlined, its shape is compared to all existing templates. The difference between a shape, $\mathbf{v}$, and the template, $\mathbf{t}$, is computed with the Manhattan error metric:

$$\mathbf{e}^{\mathbf{v}} = \sum_{i=1}^{n} \left| r_i^{\mathbf{v}} - r_i^{\mathbf{t}} \right|$$

where $r_i^{\mathbf{v}}$ is a radius in the shape $\mathbf{v}$ and $r_i^{\mathbf{t}}$ is the corresponding radius in the template $\mathbf{t}$.

To avoid simple orientation differences, the new shape is compared to the template at 16 different rotations. For each rotation angle, the size of the shape is normalized based on the length of the base radial segment. The final difference is defined to be the minimum of these errors. Reflection invariance is achieved by considering the reverse feature, $(r_n, r_{n-1}, \cdots, r_1)$, of the new shape.

Figure 3 shows the clustering process graphically. First, we set a similarity threshold, R, to a user-adjustable constant. Next, the following procedure is repeated to find a cluster. After calculating the distance between a shape and templates, the best matching template having the minimum distance can be found. If the difference between a shape and the best matching templates is below R, the shape is added to the cluster of shapes represented by the best matching template. This is done using a Point Distribution Model (PDM) [4], a statistical shape model which is generated from a set of examples of the objects to be modeled. Given a set of N shapes, $\mathbf{v}^k = \left(r_i^k\right)_{i=1\cdots n}^{k=1\cdots N}$, the average shape, $\bar{\mathbf{v}}$, is calculated using

$$\bar{\mathbf{v}} = (\overline{r_i})_{i=1}^{n} = \left(\frac{1}{N}\sum_{k=1}^{N} r_i^k\right)_{i=1}^{n}.$$

The average shape, $\bar{\mathbf{v}}$, is used as a template, $\mathbf{t}$, and updated using the new shape, $\mathbf{v}$, to make a new flexible template, $\bar{\mathbf{v}}'$,

$$\bar{\mathbf{v}}' = \frac{1}{N+1}(\bar{\mathbf{v}} \times N + \mathbf{v}).$$

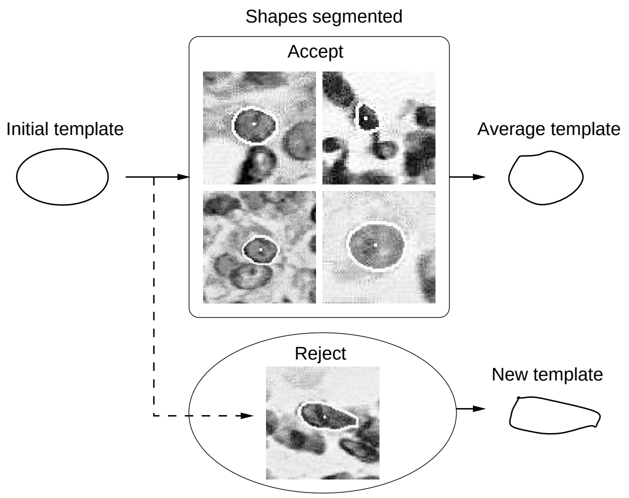


Figure 3: The clustering of shapes

If the minimum difference between a shape and any of the templates is greater than R, a new template is created with the shape of the new object. Thus, after the first image, each template represents a cluster of objects that are nearby one another in the space of possible shapes.

3.4. Classifying Objects

To solve classification problems, we use size and shape information already gathered in the clusters.

The classification method proceeds as follows. As objects are isolated by the user, they are given a class label. This requires a training phase in which an expert user is available. A count is maintained of the number of objects from each class that are represented in each cluster. These counts are stored based on the scale factor; for instance, Table 1 shows the scales and counts for the first five templates in Figure 5 for experiment 1. There are two numbers in each entry: the first one is the number of benign cells matching the template, and the second one is the number of malignant cells.

Table 1: Benign/Malignant counts for various template scales and shapes

Scale / Shape	1.0	1.2	1.4	1.6	1.8	2.0
T1	4/0	10/1	19/4	0/0	0/14	0/5
T2	4/0	29/0	19/18	0/0	0/12	0/8
T3	6/1	28/8	5/16	0/0	0/8	0/2
T4	0/3	0/4	0/1	0/0	0/1	0/0
T5	3/8	0/6	0/0	1/1	0/0	0/0

As the training proceeds, the system is able to classify new objects based on the majority class of the matching template. We saved counts by shape, size and the corresponding class. As table 1 shows, size is the more important factor than shape for cell classification. However, it is very flexible depending on applications. This simple instance-based learning scheme further improves the speed of the detection algorithm in cases where only a particular class of object is desired. For instance, if a user wants only malignant cells, the template search procedure will order the templates based on the probability of malignancy, rather than on the raw count of matched cells. Templates that match mostly benign cells will not be used until later in the search, even if their shapes were more common in previous images.

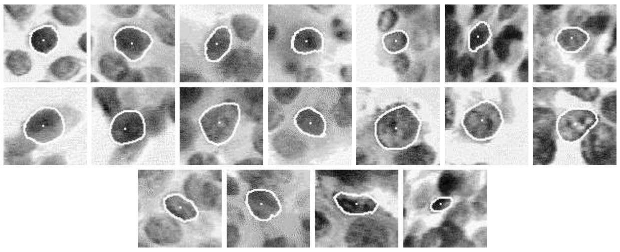


Figure 4: Classification of the breast cancer nuclei. The cells in each row belong to the same class: Benign, Malignant, and Undefined

Figure 4 shows examples of cell segmentation and classification. The fifth cell in the second row is segmented by T2 with size 1.8 and so is classified to a malignant cell according to Table 1. If the matching template has no trained instances, for example T3 with size 1.6, or has the same number of instances for multiple classes, for example T5 with size 1.6, we consider the object to be "Undefined". After a few training images in our experiments, the "Undefined" class happens rarely.

4. Experimental Results

4.1. Cytological images

The algorithm was tested on cytological images from fine-needle biopsies of breast masses, the same images used to train the original Xcyt system.[1] The images are gray-scaled with 640×480 pixel spatial resolution. These images are classified as benign or malignant on a per-sample basis; no classification is available for individual cell nuclei. Therefore in these results we consider all nuclei in benign images to be benign and all nuclei in malignant images to be malignant. This assumption is reasonable but not entirely accurate, making the classification problem particularly difficult due to classification noise.

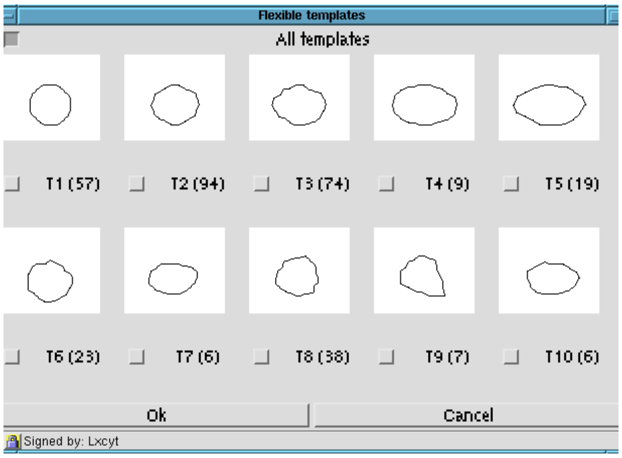
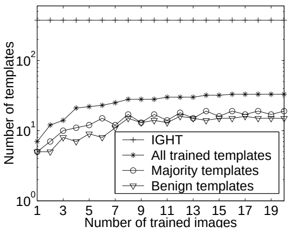


Figure 5: Learned set of cell templates

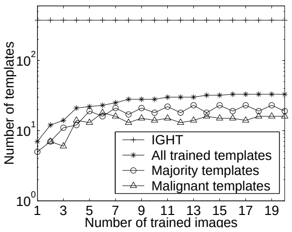
Figure 5 shows the set of templates constructed after training 333 cells on twenty images. The number next to each template name is the number of cells matching that template with various scale factors.

The system was trained on a sequence of training images, alternating between benign and malignant. Fig-

[1] Some of these images are available at `ftp://dollar.biz.uiowa.edu/pub/street/images/`.



(a) Figure 7(a)



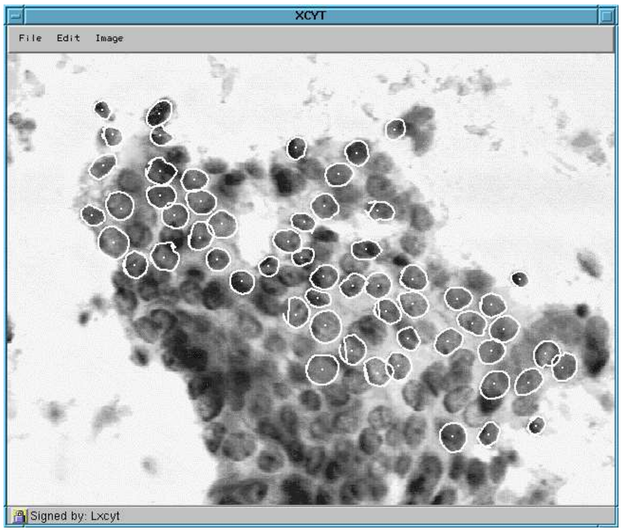
(b) Figure 7(b)

Figure 6: Number of templates searched

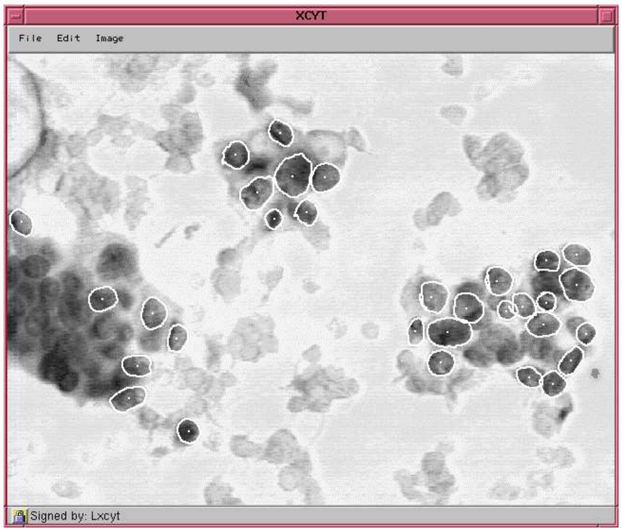
ure 6 shows the number of templates that had to be searched on a logarithmic scale. The search was carried out in four different ways. IGHT was performed with a constant set of elliptical templates. These vary significantly in size and shape in an attempt to capture the basic shape of all nuclei that might be encountered. The learning system builds a set of templates as it is used; the curve labelled "All trained templates" shows the number of these trained shapes in Figure 5 at all possible scales. After an initial ramping up period, these numbers grow very slowly at a level approximately an order of magnitude below the original set of templates.

The curve labelled "Majority templates" shows the result of searching the templates in order of their frequency in previous images. After sorting the set of templates depending on the search method, we applied 5 templates at once. Whenever the result degrades, next 5 templates are used. The search is stopped when a result equivalent in accuracy to searching all the learned templates is reached. This results in a 50% reduction in the number of necessary templates. Finally, the templates were ordered based on their classification, e.g., templates with the highest probability of being benign were searched first on the benign test image. This further reduces the necessary search time, indicating that the instance-based classification method is able to distinguish reasonably well between nuclei types, even though the ground truth classifications in this problem were extremely noisy. This bodes well for future applications on heterogeneous tissue in which a trained expert will be available for the training phase. In addition, by choosing specific templates in Figure 5, the user can reduce search time even further.

Segmentation performance was measured as follows. Two test images (one benign, Figure 7(a), and one malignant, Figure 7(b)) were set aside and tested for segmenting and classifying automatically with all trained templates after training every image. We used two measures to evaluate the segmentation performance



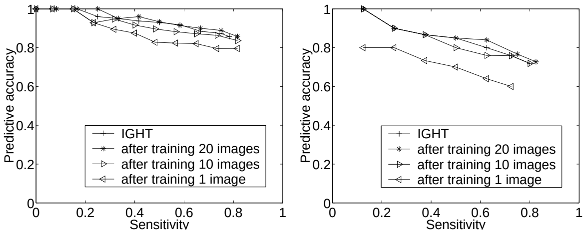
(a) benign image



(b) malignant image

Figure 7: Test images (reduced by 50%) with all trained templates

of the algorithm: sensitivity and predictive accuracy. Sensitivity is the likelihood that a nucleus will be detected in case it has been marked by an expert. Predictive accuracy is defined as the likelihood that a template match is actually associated with a nucleus segmented by an expert. Figure 8 shows that as the system is trained on more and more images, its ability to correctly segment the nuclei in the test images increases. In short, after a brief training phase, the new



(a) Figure 7(a) (b) Figure 7(b)

Figure 8: Sensitivity and predictive accuracy for segmentation with all trained templates for experiment 1

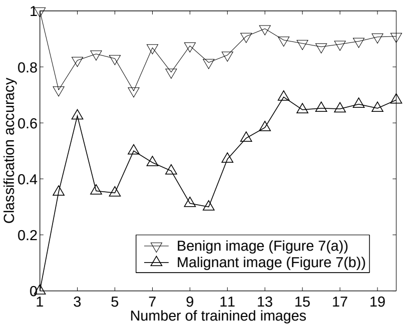


Figure 9: Classification accuracy for experiment 1

system has comparable accuracy to the original system, and achieves this performance in an order of magnitude less time per image.

We tested 10 more images, 5 benign images and 5 malignant images. Since it is typically required to segment a few dozen cells for diagnosis and prognosis, we use the predictive accuracy of the segmentation method on the first 50 outlined nuclei (in the case of the benign images) or at 80% sensitivity (in the case of the malignant images). For 408 more cells in 10 images, 250 cells in 5 benign images and 158 cells in 5 malignant images, with all trained templates, we achieved 78.19% as an average segmentation correctness: 80.4% in the benign cells and 74.68% in the malignant cells.

Table 2: Classification counts for well segmented cells

True Class	Predicted Class	
	Benign	Malignant
Benign	175	26
Malignant	28	90

Figure 9 shows that as the system is trained on more and more images, its ability to correctly classify the nuclei in the test images increases. After a few images, the classification rate stabilizes at 88% for the benign image and 65% for the malignant image. Table 2 shows the classification counts for well-segmented cells using the Table 1 templates obtained from 20 training images. For 319 cells from the segmentation test of 10 images, we achieved 83.07% as an average classification correctness: 87.06% in the 201 benign cells and 76.27% in the 118 malignant cells.

4.2. Vehicle images

The goal of this experiment is to extend the domain of the proposed system. The system for finding vehicles was trained on 25 images which are color-scaled. These images are classified as automobile, military tank, pickup track, van, and station wagon and each class has 5 training images.

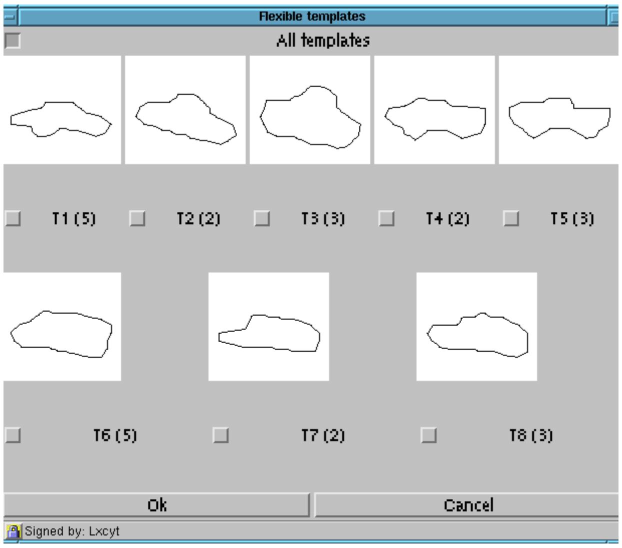


Figure 10: Learned set of vehicle templates

Figure 10 shows the set of vehicle templates, starting with no knowledge of the shapes. After training 25 images, there are two templates for tank (templates 2 and 3), pickup trucks (4 and 5) and station wagons (7 and 8). Since there were no heterogeneous templates, classification is unambiguous for all detected objects.

Figure 11 shows the segmentation results from our 15 test images. The outlines of the vehicles were accurately identified and 13 of the 15 vehicles were correctly classified. The fourth middle image was mistaken for a station wagon, because the contrast along the front side of the van is sharp and it forms a corner. The fifth bottom image was classified as a van because the template matched the shadow at the bottom of the station wagon instead of the edge along the bottom body

Figure 11: Segmentation and classification of vehicle image

of the van. The bottom row of Figure 11 shows our system has reflection invariance.

5. Conclusions and future work

This paper introduces an incremental learning scheme that aids the detection, segmentation and classification of objects in large-scale image analysis tasks. The system demonstrates dramatic improvements in the time needed to perform template-matching when compared to exhaustive search, without sacrificing the quality of the result. The system can be initialized online, removing the requirement for a pre-defined set of templates. Finally, our preliminary results show that the size and shape features are sufficient for accurate object classification.

This learning approach was chosen for its simplicity and extensibility. Its primary function in this phase is to guide the template search. However, it is certainly plausible that future applications will require other features, such as chromaticity and second-order size and shape features, to classify the objects sufficiently well. These features can also be collected and stored for each cluster. At that point, a second classification method, such as an artificial neural network, could be added to include these features in the classification.

Acknowledgments

This work was partially supported by NIH grant CA64339-04 and NSF grant IIS-99-96044. The authors would like to thank Dr. William Wolberg of University of Wisconsin Department of Surgery for his helpful suggestions and for providing the test images for the medical task.

References

[1] D. H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[2] A. Chakraborty, L. H. Staib, and J. S. Duncan. Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Transactions on Medical Imaging*, 15(6):859–870, 1996.

[3] C. C. Chang, S. M. Hwang, and D. J. Buehrer. A shape recognition scheme based on relative distances of feature points from the centroid. *Pattern Recognition*, 24(11):1053–1063, 1991.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] B. Draper. Learning control strategies for object recognition. In K. Ikeuchi and M. Veloso, editors, *Symbolic Visual Learning*, pages 49–76. Oxford Press, 1997.

[6] M.-P. Dubuisson, S. Lakshmanan, and A. K. Jain. Vehicle segmentation and classification using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3):293–308, 1996.

[7] N. Duta, A. K.Jain, and M.-P. Dubuisson-Jolly. Learning 2D shape models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 8–14, 1999.

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[9] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas. Fast and effective retrieval of medical tumor shapes. *IEEE Transactions on Knowledge and Data Engineering*, 10(6):889–904, 1998.

[10] K.-M. Lee and W. N. Street. A fast and robust approach for automated segmentation of breast cancer nuclei. In *Proceedings of the IASTED International Conference on Computer Graphics and Imaging*, pages 42–47, 1999.

[11] K.-M. Lee and W. N. Street. Dynamic learning of shapes for automatic object recognition. In *Proceedings of the 17th ICML-2000 Workshop on Machine Learning of Spatial Knowledge*, pages 44–49, 2000.

[12] M. A. Maloof, P. Langley, T. Binford, and R. Nevatia. Generalizing over aspect and location for rooftop detection. In *Proceedings of the 4th IEEE Workshop of Applications of Computer Vision*, pages 194–199, 1998.

[13] O. L. Mangasarian, W. N. Street, and W. H. Wolberg. Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43(4):570–576, 1995.

[14] H. Oh. A memory-efficient generalized Hough transform for segmenting cytological images. Master's thesis, 1997.

[15] N. Shareef, D. L. Wang, and R. Yagel. Segmentation of medical images using LEGION. *IEEE Transactions on Medical Imaging*, 18(1):74–91, 1999.

[16] W. N. Street. Xcyt: A system for remote cytological diagnosis and prognosis of breast cancer. In L. C. Jain, editor, *Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis*. CRC Press, 2000. In press.