A PIXEL CLASSIFICATION SYSTEM FOR SEGMENTING BIOMEDICAL IMAGES USING INTENSITY NEIGHBORHOODS AND DIMENSION REDUCTION

Cheng Chen[1], *John A. Ozolek*[3], *Wei Wang*[1], *Gustavo K. Rohde*[1,2]

[1]Center for Bioimage Informatics, Department of Biomedical Engineering
[2]Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA. 15213
[3] Department of Pathology, Children's Hospital of Pittsburgh, Pittsburgh, PA, 15224

ABSTRACT

We present an intensity neighborhood-based system for segmenting arbitrary biomedical image datasets using supervised learning. Because neighborhood methods are often associated with high-dimensional feature vectors, we explore a Principal Component Analysis (PCA) based method to reduce the dimensionality (and provide computational savings) of each neighborhood. Our results show that the system can accurately segment data in three applications: tissue segmentation from brain MR data, and histopathological images, and nuclei segmentation from fluorescence images. Our results also show that the dimension reduction method we described improves computational efficiency while maintaining similar accuracy.

Index Terms— pixel classification, image segmentation, intensity neighborhood, dimension reduction

1. INTRODUCTION

Automated image segmentation is becoming increasingly important in biomedical research and clinical applications. In experimental biology and medicine, tissues, cells or organs must be segmented and isolated for subsequent quantitative analysis. Due to the vast increase in the capability of image acquisition of modern devices, automated methods are often needed to segment such large amount of data. Several image segmentation methods have been described in the past (see for example [1, 2]). However, a common drawback of these methods is that most of them are not general and are specifically designed for one type of image. Thus, without a significant amount of tuning and calibration, these algorithms often cannot be employed in other applications if accurate results are necessary.

To overcome this drawback, we describe an automated segmentation system for general application in biomedical image datasets. We consider the strategy of pixel level classification utilizing supervised learning algorithms as an attractive choice because, given enough training samples (already segmented images), this strategy can be used to construct algorithms capable of performing accurately across different image modalities and different structures of tissues. We note that several works in the area have been described (see for example [3]). The majority of these, however, are based on the design (or availability) of specific numerical features that must be carefully chosen for each application. In order to make this type of system more general, we choose a nonparametric feature vector consisting of neighborhood intensities.

In our system, a few pre-segmented images are taken as input. An approach for selecting appropriate pixels (and their neighborhoods) for training the classifier is described. Our method models both variations in scale and rotation as part of the classification training procedure. To overcome the difficulties of working with high dimensional data (i.e. the number of pixels in each neighborhood window) we also describe a method based on principal component analysis (PCA) to reduce the dimensionality of the data, improving the computational efficiency while maintaining accurate results. We apply this segmentation system to three different datasets: brain tissue segmentation from magnetic resonance (MR) image data, segmentation of tissues in histopathological images, and segmentation of nuclei from fluorescence microscopy images. We compare our results to the results produced by state of the art methods for each application. Our results show that our system can achieve comparable or better accuracy than some state of the art methods in each application. Our results also show that the dimension reduction method improves computational efficiency with comparable results.

2. METHODS

As mentioned above, our supervised learning-based segmentation system takes as input a few pre-segmented example images and for all the pixels in the example images, each labeled pixel's square neighborhood is utilized and reordered into a nonparametric feature vector for that pixel. For this idea to work well across different applications, the information content provided by the example images must be mined carefully. It is important to model intensity differences that may occur between different images of the same modality, as well as variations in scales and orientations that may occur. Empirical approaches for mining such information are described in sequence below. With such an approach, however, even a few pre-segmented images could generate several million pixels to be available for training a classifier. The computational cost of training a classifier with P training examples usually ranges from P^2 to P^3 [4]. To overcome this limitation, we will investigate two aspects: (1) reducing the amount of training data by selecting "important" pixels into the training set; and (2) reducing the dimensionality of feature vectors.

Our system consists of a training stage and a testing stage. In the training stage, we normalize the input data, model scaling and orientations of the input data via filtering and resampling, and extract important pixel windows while reducing their dimensionality. We then train several classifiers at different scales. In the testing stage, predictions for new unknown data are made by combining (via a voting procedure) the trained classifiers at different scales.

2.1. Data Normalization

Intensity normalization for input data is important because variations in intensity from image to image may prevent high classification accuracy. Many papers have been published for solving specific normalization problems in different modalities of imaging. Common examples include: correction of different bias fields in MRI, illumination heterogeneity in fluorescence imaging, etc. In this specific implementation of our system, we opt for a simple approach where all pixel intensities in the training and testing images are uniformly mapped to the [0,1] range, while the top (and bottom) 1% pixels are discarded. This approach is used in all experiments below. We note that in the presence of application specific information, the intensity normalization procedure can be improved.

2.2. Window selection and rotational variation modeling

For a set of labeled pixels, each pixel's neighborhood intensities are selected as a general nonparametric feature vector. Neighborhood intensities are constructed by imposing a pre-defined size window centered at the given pixel, in which pixels inside the window are regarded as neighbors. In this paper, we choose square windows of size $N \times N$ for 2D data, and cube window of size $N \times N \times N$ for 3D data (different shapes of windows and their performances will be explored in future work), centered at the given pixel (N is an odd number). Pixels' intensities inside the window are chosen and reordered to comprise a N^2 (or N^3 in 3D) nonparametric feature vector associated with the given pixel. In addition, we note that the intensity neighborhood vectors comprised from square or cube windows are not rotationally invariant. In most biomedical imaging applications of interest, however, a fixed coordinate frame cannot be assumed and rotational variation must be accounted for. We therefore virtually augment the set of training pixels to include rotated versions of each window. This is accomplished by rotating each example image using linear interpolation. In addition, we include flipped (coordinate reversed) versions of windows to increase useful information. Naturally, computational complexity constraints limits us to model only a finite number of rotations and flips, and the details are described in Section 3.

2.3. Modeling scale information

We also note that due to different structure sizes of object tissues, organs, or cells, information at different scales is often important for appropriate classification. We extract information at different scales from the images by utilizing a standard multi-resolution procedure [5] in which the training images are first convolved with a Gaussian kernel for smoothing at each scale$s \in \{0, 1, 2, ..., S\}$, and the neighborhood patches are assembled by subsampling at every 2^s pixels to construct the neighborhood vectors.

2.4. Reducing the size of the training set

As mentioned above, even a few example labeled images (especially when scales and variations are modeled explicitly as above) can generate several millions of training pixels. We therefore describe an empirical procedure to select "important" neighborhoods from the entire set of labeled neighborhood (including those artificially rotated, flipped and scaled as described above). We first define the pixel near the boundaries of two or more classes as boundary-type pixel and the pixel inside the region of single class as interior-type pixel, and we assume that it is more important to include rotated versions of neighborhoods of boundary-type pixel. Therefore, we only augment the set of windows by rotating and flipping pixels whose window neighborhood contains pixels that belong to two or more classes. We also assume that much redundant information is contained in each class (type of tissue to be segmented). Therefore, we implement the K-means based algorithm for selecting only a set of Q representative neighborhoods in the following steps:

1. A K-means procedure is used to divide the input labeled training images into R non-overlapping spatial regions (R is manually defined).
2. Each of the pixels in the regions defined in step 1 are categorized as a boundary-type or interior-type pixel. In each region, the K-means procedure is used to select only certain pixels' intensity neighborhoods for training (the amount selected is calculated based on R and Q). This is done for both boundary (including rotated and flipped windows) and interior type pixels, and is repeated for each scale s.
3. Finally, for each scale s, all the clustered samples for both boundary-type and interior-type of all subsets and all classes are combined as the training set containing Q samples and thus, several training sets associating with different scales $s \in \{0, 1, ..., S\}$ can be built.

2.5. Reducing the dimensionality of the training set

In addition to the procedure for reducing the number of training pixels explained above, we also utilize a PCA-based procedure to reduce the size of each feature vector (which in our case is composed of intensity neighborhoods). Since neighborhood feature vector is typically high dimensional, for example, for a $N \times N \times N$ neighborhood, the dimensionality becomes N^3, which will increase greatly when the neighborhood size increases. Thus, we consider reducing the dimensionality of vectors by first converting the intensity neighborhood vector to a feature vector containing uncorrelated variables via principal component analysis (PCA) [6] and then discard useless variables for classification. Based on an existing training set containing Q samples of N^3 dimensions, PCA method is implemented in a standard procedure in which a covariance matrix is built from the samples and the eigenvectors of the covariance matrix are calculated and sorted in order of descending eigenvalues; then, for each sample in the training set, the dot product between each eigenvector and that pixel's neighborhood vector is calculated as the coefficient in this sample's PCA based feature vector. Then, stepwise discriminant analysis (SDA) [6] is used for automated determination of subspace dimensionality and selection of the principal components from the PCA feature vectors. The SDA method chooses the variables from the feature vector which maximize the ratio of the variance between the classes to the variance within the classes. Thus, we are able to produce a new training set with the same samples but associating with a lower dimensional PCA based subspace feature vector by selecting the necessary principal components from the PCA based feature vector using SDA method.

Additionally, we note that the computational complexity of actual classification (testing phase) using Support Vector Machine (SVM) with a non-linear kernel such as RBF kernel is $O(nd)$, where n denotes the number of support vectors, and d the feature dimensions [7]. Thus, based on the same training set but with a lower dimension (PCA vector), computational efficiency can be improved.

2.6. Training Stage and Testing Stage

In our current implementation, one classifier is trained for each image scale (computed as described above). We utilize the support

vector machine (SVM) classifier, in which the Gaussian radial basis function (RBF) kernel is chosen, and parameters such as weights, RBF kernel size, etc. are optimally selected by the cross validation procedure described in [7]. For multi-class classification, we use the "one-against-one" approach [8], and the popular "LibSVM" software [7] is used for implementation.

In the testing stage, the final result is predicted through a confidence voting strategy among the results from individual classifiers trained at each scale. For a given pixel i, the prediction label provided by the individual classifier at scale s ($s \in 0, 1, ..., S$) is denoted as L_i^s, and the confidence of assigning class c ($c \in 1, 2, ..., C$) to the pixel i at scale s is denoted as $F(L_i^s = c)$. The final prediction label L_i^{final} is calculated as follows:

$$L_i^{final} = \underset{c \in \{1,2,...,C\}}{\arg\max} \sum_{s=0}^{S} F(L_i^s = c) \quad (1)$$

In our work, we choose the posterior probability $Pr(L_i^s = c/x_i)$ (x_i is the intensity neighborhood vector/ PCA based subspace feature vector for the given pixel i) as the confidence, and the multi-class posterior probabilities in SVM can be estimated by combining all the pairwise class probabilities [9].

3. EXPERIMENTAL EVALUATION

In this section, we test our system in three different applications. In addition, we explore the effect of dimensionality reduction using the PCA-based approach described above. We evaluate the performances both qualitatively and quantitatively and compare our results to those produced by state-of-the-art algorithms specifically designed for each application. We choose the neighborhood size to be $3 \times 3/3 \times 3 \times 3$ for 2D/3D data (performances of different neighborhood sizes will be explored in future work). When required, images are rotated along the Z axis both clockwise and counter-clockwise by 45 degrees. They are flipped left and right, up and down in X-Y plane for each orientation so a total 9 neighborhood variations under totally 5 scales (scale s from 0 to 4). We also note that these parameters (the parameters pertaining to the general segmentation method we propose) are keep constant for all the experiments in this paper. The segmentation tasks and datasets to which we apply our algorithm are:

- *Segmentation of brain MR images:* The IBSR real T1-weighted brain-MR dataset [10] is chosen, in which three tissues, cerebrospinal (CSF), gray matter (GM), and white matter (WM) are to be segmented. Each brain data contains 128 scanned slices. The dataset makes available expert hand segmentations of each brain in the dataset, which we use for training and testing the classifier. In our tests, 4 brain images were randomly selected for training, and 14 brain images were used for testing. In order to evaluate the quality of our segmentations, we compute the dice metric for the 14 test datasets [11]. Results are compared to the work of Awate et al [11] who showed overall high quality in segmenting the same dataset using a semi-supervised learning approach.
- *Segmentation of H&E stained histopathology images of teratomas* [3]: we utilize a dataset where the labeled data was provided by a pathologist (JAO). We note that these are color (RGB) images. This is a difficult problem since teratomas are notorious for not having well known tissue organization. We focus on segmenting four tissue types: background/other tissues (O), bone (B), cartilage (C) and fat (F). 4 images were randomly selected for training and 10 images were randomly selected for testing. Since there is no current robust method for tissue segmentation in teratoma histology images, we choose the popular color K-means method [12] for comparison. Mean classification accuracy is used to assess accuracy.
- *Segmentation of nuclei from fluorescence images:* expert segmented nuclei images are available from Dr. Murphy's group at Carnegie Mellon University [13]. The data consists of 48 2D images, out of which 6 images were randomly selected for training and the remaining images were used for testing. The results were assessed following Coelho et al. [13]. Three types of metrics are used for evaluation: (1) the Rand and Jaccard indices, which measure the fraction of the pixel pairs where the segmentation and the ground truth agree. The larger the value, the better the result. (2) Error Counting: split, merged, added, and missing, which counts the errors in segmenting nuclei. (3) Spatially-Aware Evaluation: NSD and Hausdorff metrics, which measure the distance of segmented nuclei to the ground truth. We calculate the mean values for all test images, and compare with the best results in [13].

3.1. Results

Figure 1 contains raw unprocessed images (a,d,g), ground truth (c,f,j), and results of using our segmentation method without dimension reduction (h) and with dimension reduction (b,e,i). We note that the results presented here are without any post processing operations. In our experience, results for each application can be made visually more appealing (and quantitatively more accurate) by utilizing simple post-processing operations such as morphological closing and opening. Table 1 contains quantitative results on each application as described above. In the brain MR image dataset, on average, our method (both non-PCA and PCA) performs as well or better than the method described by Awate et al [11]. In the teratoma histology dataset, the overall classification accuracy of four tissues (both non-PCA and PCA) outperforms the color K-means method. In the nuclear dataset, although Hausdorff distance metric and added error metric of our results are much larger than the compared result due to the noise effect, which can be easily improved using some simple morphological operations, the overall performances are comparable to the results in [13]. We note that the PCA results are, on average, only slightly less accurate than the non-PCA results.

4. CONCLUSION AND DISCUSSION

In this paper, we proposed a supervised learning based system for segmenting biomedical images. The system is based on a neighborhood intensity window approach for characterizing each pixel, taking into account rotational and flipped variations, as well as multiple scales. An empirical procedure for selecting a sub-portion of training pixels, together with dimension reduction operation, was described.

The method was applied to three different biomedical image segmentation tasks: segmentation of brain tissues from MR images, segmentation of teratoma tissues from histopathology images, and segmentation of nuclei from fluorescence microscopy images. Results showed that the segmentation produced by our methods compared favorably (on average as well or better) to other comparable methods. In addition, the results showed that when reducing the dimensionality from 27 (non-PCA vector) to 14 (PCA vector), for example, the average practical computational time for training a

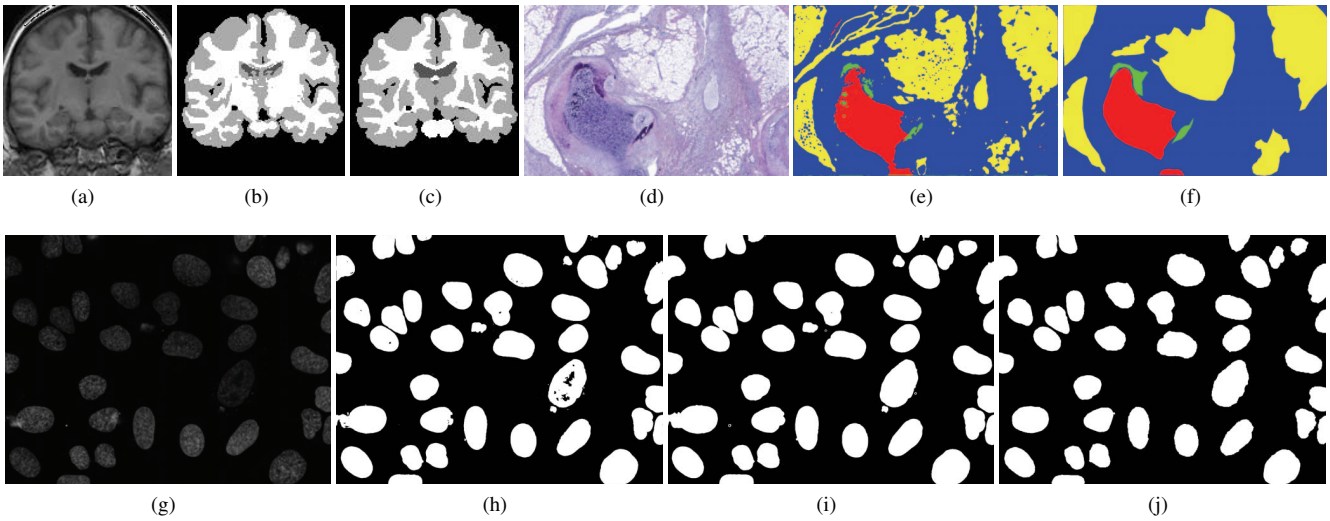

Fig. 1. Performance of our system in three applications: (a), (d) and (g) are original brain-MR, histology and nuclei images, respectively. Panels (b), (e) and (i) are results produced by our system with dimension reduction. Panels (c), (f) and (j) are ground truth images. In addition, we show the nuclei result using non-PCA vectors in Panel (h).

Table 1. Quantitative Evaluation for Three Applications

Case	Non-PCA	PCA	Comparison
brain (GM)	0.9028	0.8953	0.8074
brain (WM)	0.8185	0.8024	0.8868
histology (O)	89.39%	86.19%	55.20%
histology (B)	53.01%	29.03%	29.79%
histology (C)	69.11%	70.45%	51.06%
histology (F)	85.76%	85.08%	58.73%
nuclei (RI)	96%	96%	96%
nuclei (JI)	2.5	2.5	2.2
nuclei (Hausdorff)	130.2	132.6	12.9
nuclei (NSD×10)	0.86	1.12	0.7
nuclei (Split)	0.8	0.2	1.1
nuclei (Merged)	2.8	3.0	1.2
nuclei (Added)	6.1	5.8	0.3
nuclei (Missing)	0.1	0.1	2.9

SVM model for brain segmentation reduces from 4.48×10^4 seconds to 1.53×10^4 seconds (approximately 66% savings) and the time for segmenting one brain data reduces from 1.14×10^4 seconds to 9.33×10^3 seconds (approximately 18% savings), which proves the computational efficiency of using PCA vectors. Finally, we note that the comparable methods were often designed for each specific application, while our system is general: the same system, with identical training procedure, was used for all applications.

5. REFERENCES

[1] N. Malpica, C.O. de Solórzano, J.J. Vaquero, A. Santos, I. Vallcorba, J.M. García-Sagredo, and F. del Pozo, "Applying watershed algorithms to the segmentation of clustered nuclei," *Cytometry Part A*, vol. 28, no. 4, pp. 289–297, 1997.

[2] J.A. Sethian et al., "Level set methods and fast marching methods," *Journal of Computing and Information Technology*, vol. 11, no. 1, pp. 1–2, 2003.

[3] A. Chebira, J.A. Ozolek, C.A. Castro, et al., "Multiresolution identification of germ layer components in teratomas derived from human and nonhuman primate embryonic stem cells," in *Proc. IEEE Int. Symp. Biomed. Imaging*, 2008, pp. 979–982.

[4] C.M. Bishop et al., *Pattern recognition and machine learning*, Springer New York:, 2006.

[5] T. Lindeberg, "Scale-space for discrete signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 234–254, 1990.

[6] T. Hastie, R. Tibshirani, and J.H. Friedman, *The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations*, Springer New York:, 2001.

[7] C.C. Chang and C.J. Lin, "Libsvm: a library for support vector machines," 2001.

[8] S. Knerr, L. Personnaz, G. Dreyfus, and et al., "Single-layer learning revisited: A stepwise procedure for building and training a neural network," *Optimization Methods and Software*, vol. 1, pp. 23–34, 1990.

[9] T.F. Wu, C.J. Lin, and R.C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *The Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[10] Internet Brain Segmentation Respository (IBSR), "http://www.cma.mgh.harvard.edu/ibsr," .

[11] S.P. Awate, T. Tasdizen, N. Foster, and R.T. Whitaker, "Adaptive, nonparametric markov modeling for unsupervised, mri brain-tissue classification," *Medical Image Analysis*, vol. 10, no. 5, pp. 726–739, 2006.

[12] MATLAB Image Processing ToolBox, "Color-based segmentation using k-means clustering," .

[13] L.P. Coelho, A. Shariff, and R.F. Murphy, "Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms," in *ISBI 2009*. IEEE, 2009, pp. 518–521.