Automatic Segmentation using Learning Algorithms for High-Resolution Magnetic Resonance Imaging of the Larynx

Reeve Ingle – ringle@stanford.edu
Kie Tae Kwon – ktkwon07@stanford.edu

1. Introduction

Laryngeal cancer is one of the most common types of head and neck cancer. Depending on the stage of the tumor, chemotherapy, partial laryngectomy, or total laryngectomy may be required. Accurate staging of the tumor is necessary to properly treat laryngeal cancer and avoid unnecessary procedures such as total laryngectomy, which can significantly degrade a patient's quality of life [1].

Accurate staging of laryngeal cancer can be clinically challenging due to the difficulty of detecting the extent of laryngeal cartilage invasion by the tumor. To resolve this limitation, high-resolution Magnetic Resonance Imaging (MRI) of the larynx [2] has been investigated and has led to the availability of high-resolution 3D MRI datasets with multiple contrasts. These datasets are well-suited for automatic image segmentation of the larynx (i.e., classification of physiological structures and tissues). While fully automated segmentation of the laryngeal cartilages remains unexplored, a multi-contrast and multi-dimensional approach has proven useful for segmenting articular cartilage [3].

The purpose of this project was to investigate the application of learning algorithms to automatically segment high-resolution MR images of the larynx, which can potentially increase the accuracy of laryngeal cancer staging.

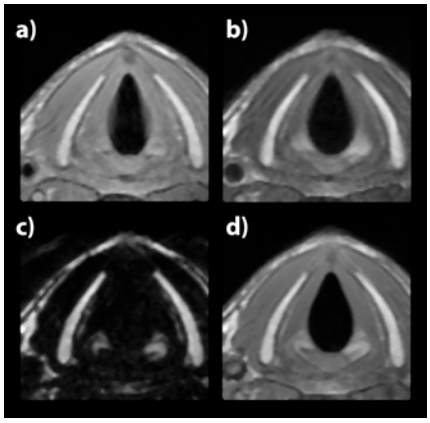

Figure 1. Images from a single slice of a larynx dataset. Four MR sequences were chosen to yield different levels of contrast: PD (a), SE (b), FSE-IDEAL (c), and FSE-XL (d).

2. Methods

In this project, we applied learning algorithms to implement automatic segmentation of MR images of the larynx. Among the various tissues in the larynx, our focus was on the laryngeal cartilage. Using the different T1 and T2 relaxation times of cartilage versus surrounding tissues, we selected appropriate MR sequences that produced images with different contrast levels [4]. Each pixel in the resulting dataset was represented as a vector of different intensity levels. Figure 1 illustrates the different contrast levels produced by the four different MR sequences that were used to image the larynx of healthy volunteers. Both a supervised learning algorithm (support vector machine, SVM) and an unsupervised learning algorithm (k-means) were investigated. For both algorithms, the vector-valued pixel intensities were used as features. For the SVM, a subset of larynx images was manually segmented for training and quantitative assessment of testing accuracy. The segmentation classes consisted of: 1) ossified laryngeal cartilage and fat, 2) muscle, and 3) trachea and background pixels. These regions are denoted with arrows in Fig. 2.

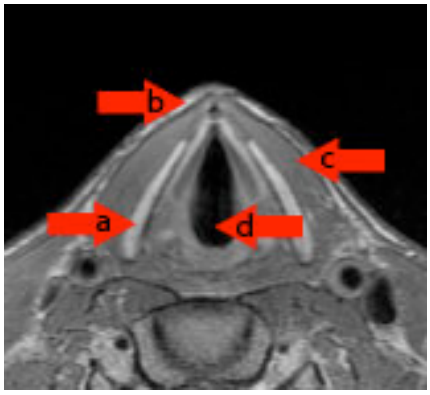

Figure 2. MR image of the larynx showing laryngeal cartilage (a), subcutaneous fat (b), muscle (c), and trachea / background (d).

2.1. Data Acquisition

A larynx-dedicated three-channel array coil [2] was used to scan two healthy volunteers on a GE 1.5 T MRI system. Four 3D MR sequences, proton-density-weighted spin-echo (PD), spin-echo (SE), fast spin-echo IDEAL (FSE-IDEAL), and fast spin-echo XL (FSE-XL), were run to acquire four sets of images. These sequences were chosen based on their ability to provide different contrast among cartilage, muscle, and other tissues of interest.

2.2. Procedure

A series of Matlab routines was implemented to perform various preprocessing steps to align images (registration), correct for the MR coil sensitivity pattern (intensity correction), and structure the data into a

format suitable for SVM and k-means. For the SVM, we used the LibSVM package [5] to implement a multi-class SVM model. For k-means, we directly implemented the algorithm in Matlab. We tested the performance of both algorithms on two larynx MR datasets. A detailed description of the procedure is given below.

2.2.1. Preprocessing

2.2.1.1. Registration

Image registration was used for spatial alignment of images to correct for subject motion that may have occurred between scan sequences. Registration was performed automatically using mrVista, a Matlab interface for analyzing functional and anatomical data [6].

2.2.1.2. Intensity Correction

Since a larynx-dedicated array was used to acquire the MR datasets, image intensity varied significantly with spatial location relative to the array. Intensity correction was performed to account for the coil sensitivity profile that made regions near the array undesirably bright [7]. We implemented an intensity correction method that fit a low-order polynomial to the proton-density images, which represent the coil sensitivity [8]. Amplitude thresholding was used to mask background and low-SNR regions of the image, and polynomial coefficients were computed by solving the following convex optimization problem

$$\begin{aligned} \min. \quad & \left\| W(Xa - y) \right\|_2^2 \\ subject\ to \quad & Xa > 0 \end{aligned} \tag{1}$$

where W is the diagonal matrix of binary mask weights, X is the regressor matrix containing powers and cross terms of x and y coordinates up to the desired fitting order, y is the vector of original image values, and a is the vector of polynomial coefficients.

2.2.2. Classification

2.2.2.1. SVM

Manual image segmentation was done using 3DSlicer, an Insight Segmentation and Registration Toolkit (ITK) based software [9]. Pixels were manually labeled as one of three classes:

1) Ossified laryngeal cartilage and fat
2) Muscle
3) Trachea and background

For training datasets, only regions that could be labeled with high confidence were used. For testing datasets, all pixels were labeled, as the manual segmentation was used for quantitative assessment of testing accuracy. Subcutaneous fat and ossified laryngeal cartilage (fatty marrow) were given the same label due to their similar contrasts in all images, making them difficult to distinguish in manual segmentation. Likewise, the trachea and background were given the same label since these regions do not produce an MR signal.

In each dataset, the MR slices covering the laryngeal cartilages were chosen for processing. Datasets from two different subjects were used for SVM training and testing. LibSVM [5] was used to implement and solve the multi-class SVM. A Gaussian kernel was selected for the SVM model.

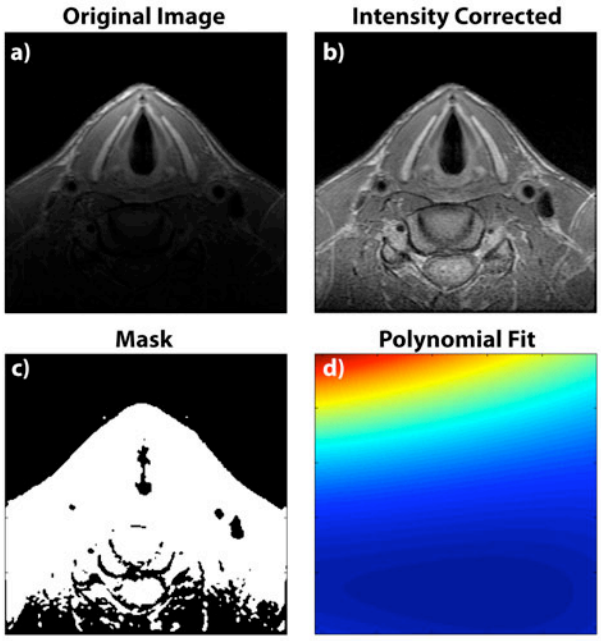

Figure 3. Intensity correction on one slice of the proton-density-weighted dataset: original (a), corrected (b), mask (c), and resulting third-order polynomial fit (d).

The SVM performance was assessed by comparing the SVM classification with the manually segmented test data. An accuracy score was computed as the ratio of correctly classified pixels to total pixels in the region of interest.

2.2.2.2. K-means

The k-means clustering algorithm was carried out on the same datasets used for SVM testing. The algorithm was run with $k=3$ clusters, which allowed direct comparison with the manually segmented testing images and SVM results. The algorithm was repeated four times, and the solution with the lowest objective value was used to avoid bad local minima.

3. Results

3.1. Intensity Correction

Figure 3 shows one slice from the testing dataset before and after intensity correction. The third-order polynomial fit to the original image resulted in good intensity correction in a computation time of less than 30 seconds on a 2.0 GHz Intel Core 2 Duo processor. Intensity correction significantly improved the performance of both SVM and k-means classification. Figure 4 shows an example of the performance of SVM and k-means with and without intensity correction. In both cases, the quality of the resulting segmentation was greatly improved by intensity correction. Specifically, intensity correction improved the overall accuracy of the SVM by 5%, the percentage of correctly labeled laryngeal cartilage pixels by 9%, and the percentage of correctly labeled muscle pixels by 5%. Confusion matrices for the SVM results are given in Tables 1 and 2.

3.2. SVM

Figure 5 shows the scatter plot of all pixels in the testing dataset, with three of the four contrast levels plotted on the x, y, and z axes. The label of each pixel was determined by manual segmentation. Each species forms a well-localized cluster, suggesting that automatic segmentation will yield good accuracy. Figure 6 shows three slices from the testing dataset. The top row shows one of the four contrasts, the second row shows the manual segmentation, and the third row shows the SVM classification. For the SVM, training and testing datasets were acquired from larynx scans of two different subjects. The SVM achieved an overall accuracy of 93%. Since there were a disproportionate number of background and muscle pixels compared to cartilage pixels, the confusion matrix was computed to give a more detailed performance metric. The confusion matrix is given in Table 3. For comparison with multi-subject training and testing, an SVM was trained and tested on different slices from the same subject. The overall accuracy was 94%, and the confusion matrix is given in Table 4. As expected, the accuracy of single-subject training and testing was better than that of multi-subject training and testing. The largest change in accuracy came from the cartilage pixels, with an increase of 8% (from 69% to 77%). Although multi-subject training and testing had a lower accuracy than single-subject training and testing, the results were quite promising considering the physiological differences among subjects, such as the degree of ossification of laryngeal cartilage, which significantly affects the intensity level of the signal. These differences can potentially be accounted for

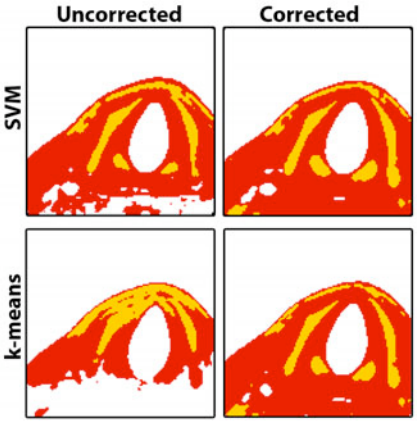

Figure 4. Results of SVM (top row) and k-means (bottom row) on an uncorrected image (left column) and an intensity-corrected image (right column). Due to the uneven coil sensitivity profile, image intensity varies with position. Intensity correction significantly improves segmentation results.

Table 1 – Confusion Matrix Uncorrected Image

	Mscl.	Cart.	Bkgd.
Mscl.	0.83	0.06	0.11
Cart.	0.25	**0.75**	0.00
Bkgd.	0.03	0.00	0.97

Table 2 – Confusion Matrix Intensity Corrected Image

	Mscl.	Cart.	Bkgd.
Mscl.	0.88	0.08	0.04
Cart.	0.16	**0.84**	0.00
Bkgd.	0.01	0.00	0.99

Table 3 – Confusion Matrix Multi-Subject Train/Test

	Mscl.	Cart.	Bkgd.
Mscl.	0.89	0.07	0.04
Cart.	0.29	**0.69**	0.02
Bkgd.	0.01	0.00	0.99

Table 4 – Confusion Matrix Single-Subject Train/Test

	Mscl.	Cart.	Bkgd.
Mscl.	0.90	0.08	0.02
Cart.	0.23	**0.77**	0.00
Bkgd.	0.01	0.00	0.99

using additional preprocessing methods such as histogram equalization.

A Gaussian kernel with parameter $g=0.25$ ($\sigma^2=2$) was used in the SVM, which is the default kernel used by LibSVM. We also tested other built-in kernels supported by LibSVM, including linear, polynomial, and sigmoid kernels. There were no significant differences between the Gaussian and linear kernels, and the overall accuracy, confusion matrices, and SVM segmentations were very similar for these kernels. For the polynomial and sigmoid kernels, the performance was similar or worse than that of the Gaussian kernel among the range of parameters we tested.

One drawback of using a supervised learning algorithm for this application is that the manual segmentation of training data can be very time consuming. Furthermore, manual segmentation is prone to errors. For example, the thin layer of subcutaneous fat can be especially challenging to manually segment due to its irregular thickness and non-contiguous structure. Because of these challenges, an unsupervised learning algorithm (k-means) was investigated since it requires no manual segmentation of images.

3.3. K-means

The results of k-means are shown in the last row of Figure 6. K-means produced comparable segmentation results as the SVM without requiring manual segmentation. Since k-means assigns class labels randomly, we manually assigned each of the three output classifications to the appropriate class (laryngeal cartilage/fat, muscle, or trachea/background). By defining each output class in this way and using the manually segmented testing images as the gold standard for accuracy, we can compute an accuracy score and confusion matrix for k-means, allowing us to quantitatively compare its performance to the SVM. The overall accuracy of k-means was 94% with respect to the manually segmented testing images. The confusion matrix is shown in Table 5. The k-means confusion matrix was almost identical to that of the single-subject SVM (i.e., training and testing data from one subject) shown in Table 4. When compared to the confusion matrix for the multi-subject SVM (Table 3), the accuracy of the laryngeal cartilage classification improved by 8%, from 69% for the SVM to 77% for k-means.

3.4. MR Sequence Selection

Each MR sequence we acquired had a scan time ranging from two to ten minutes. Due to practical limits on the total duration of the MR exam as well as the increased potential for motion (both during and between scans) from long scan times, we would like to use the fewest number of contrasts that still yields favorable accuracy, since this requires the least amount of scan time. We investigated the effects of using only a subset of the four contrasts. Since the proton density sequence was required for intensity

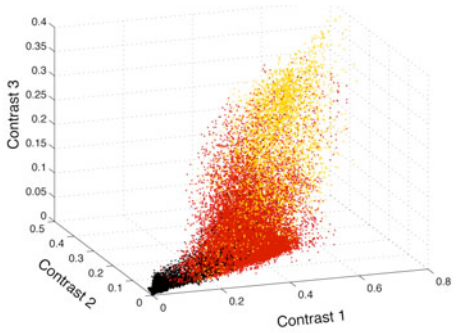

Figure 5. Scatter plot showing contrast levels of the manually segmented testing image for contrast 1 (PD), contrast 2 (SE), and contrast 3 (FSE-IDEAL). Cartilage and fat (yellow), muscle (red), and trachea and background (black) form well-localized clusters suggesting that automatic segmentation will yield good accuracy.

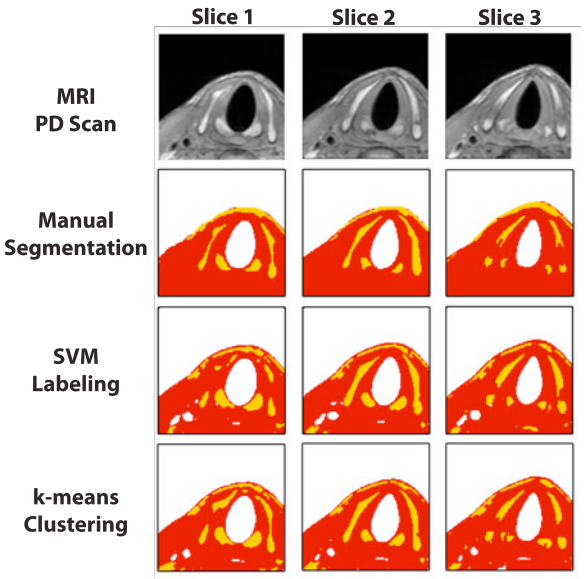

Figure 6. SVM and k-means testing results. Proton density MR images (row 1), desired labeling (row 2), SVM labeling (row 3), and k-means labeling (row 4) for three test slices from a larynx dataset. Results of SVM labeling demonstrate classification of cartilage and fat (yellow), muscle (red), and trachea and background (white) with 93% accuracy with respect to manual segmentation. K-means results in comparable segmentation results without requiring tedious manual segmentation.

Table 5 – Confusion Matrix K-means

	Mscl.	Cart.	Bkgd.
Mscl.	0.91	0.07	0.02
Cart.	0.22	**0.77**	0.01
Bkgd.	0.01	0.00	0.99

correction, this resulted in three possible contrast pairs when only two contrasts were used for SVM training and testing. Among those pairs, we found that the combination of PD and SE sequences resulted in the most accurate segmentation. Figure 7 shows the segmentation results when each pair is used for SVM training and testing. Qualitative assessment of the segmented images shows that the PD and SE sequence pair results in the best segmentation of laryngeal cartilage. Furthermore, quantitative comparison of confusion matrices shows that the accuracy of laryngeal cartilage classification is highest for this pair of sequences (79%, 64%, and 56% for PD+SE, PD+FSE-IDEAL, and PD+FSE-XL, respectively).

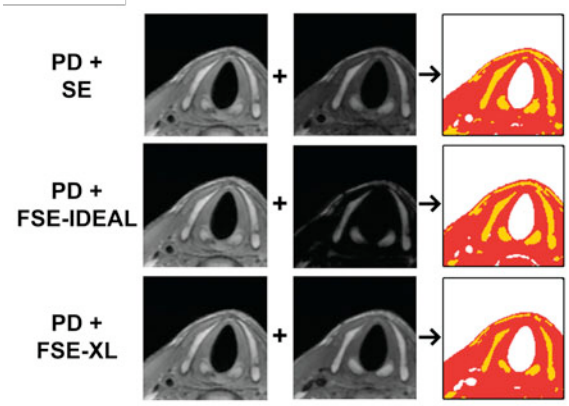

Figure 7. Comparison of SVM performance using only two contrasts for SVM training and testing. Results demonstrate improved classification when the PD and SE sequences are used (top row).

4. Conclusions

We have successfully implemented SVM and k-means algorithms to segment the cartilages from MR images of the larynx. An intensity correction technique was implemented, resulting in significant improvements in the performance of segmentation. SVM classification using multi-subject training and testing achieved an overall accuracy of 93%, muscle classification accuracy of 89%, laryngeal cartilage classification accuracy of 69%, and trachea and background accuracy of 99% with respect to manual segmentation. Unsupervised k-means produced comparable segmentation results without the need for manual segmentation, which can be tedious and error-prone.

For the SVM, future work includes the investigation of specialized kernels derived from MR physics, which can lead to better classification of the different species based on MR parameters such as T1 and T2 relaxation times. For k-means, future work includes comparison with other unsupervised learning algorithms, such as Expectation-Maximization (EM). Additional work includes the extension of this project to reconstruct an entire 3D larynx dataset, which can potentially increase the accuracy of laryngeal cancer staging.

5. Acknowledgements

We would like to thank Joëlle Barral for her guidance, direction, and expertise throughout this project. We would also like to thank Berhane Azage, whose initial work on larynx segmentation inspired this project.

6. References

[1] Forastiere A., *et al.* "Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer," N Engl J Med, 349 (22), p. 2091-2098, 2003.

[2] Barral J.K., *et al.* "High-resolution larynx imaging," Proc. of the 17th Annual Meeting of ISMRM, Honolulu, HI, 2009, p. 1318.

[3] Koo S., *et al.* "Automatic segmentation of articular cartilage from MRI: A multi-contrast and multi-dimensional approach," Proc. of the 16th Annual Meeting of ISMRM, Toronto, Canada, 2008, p. 2546.

[4] Vannier M.W., *et al.* "Validation of MRI multi-spectral tissue classification," Comput Med Imaging Graph, 15 (4), p. 217–223, 1991.

[5] Chang, C.C., *et al.* "LIBSVM: a library for support vector machines," 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[6] http://white.stanford.edu/software/

[7] Vovk U., *et al.* "A review of methods for correction of intensity inhomogeneity in MRI," IEEE Trans Med Imag, 26 (3), p. 405-421, 2007.

[8] Styner M., *et al.* "Parametric estimate of intensity inhomogeneities applied to MRI," IEEE Trans Med Imag, 19 (3), p. 153-165, 2000.

[9] Gering D., *et al.* "An integrated visualization system for surgical planning and guidance using image fusion and an open MR," J of Mag Reson Imag, 13 (6), p. 967-975, 2001.