

Automatic Anatomy Recognition in Post-Tonsillectomy MR images of obese children with OSAS

Yubing Tong¹, Jayaram K.Udupa¹, Dewey Odhner¹, Sanghun Sin², Raanan Arens²

¹Medical Image Processing Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104

²The Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY, 10467.

ABSTRACT

Automatic Anatomy Recognition (AAR) is a recently developed approach for the automatic whole body wide organ segmentation. We previously tested that methodology on image cases with some pathology where the organs were not distorted significantly. In this paper, we present an advancement of AAR to handle organs which may have been modified or resected by surgical intervention. We focus on MRI of the neck in pediatric Obstructive Sleep Apnea Syndrome (OSAS). The proposed method consists of an AAR step followed by support vector machine techniques to detect the presence/absence of organs. The AAR step employs a hierarchical organization of the organs for model building. For each organ, a fuzzy model over a population is built. The model of the body region is then described in terms of the fuzzy models and a host of other descriptors which include parent to offspring relationship estimated over the population. Organs are recognized following the organ hierarchy by using an optimal threshold based search. The SVM step subsequently checks for evidence of the presence of organs. Experimental results show that AAR techniques can be combined with machine learning strategies within the AAR recognition framework for good performance in recognizing missing organs, in our case missing tonsils in post-tonsillectomy images as well as in simulating tonsillectomy images. The previous recognition performance is maintained achieving an organ localization accuracy of within 1 voxel when the organ is actually not removed. To our knowledge, no methods have been reported to date for handling significantly deformed or missing organs, especially in neck MRI.

Keywords: multi-object segmentation, missing organs detection, Automatic Anatomy Recognition-AAR, Adeno-tonsillectomy

1. INTRODUCTION

Childhood obesity has more than doubled in children and quadrupled in adolescents in the past 30 years [1]. Pediatric Obstructive Sleep Apnea Syndrome (OSAS) is one of the most detrimental effects of childhood obesity. OSAS is also known to introduce additional health issues in children including neurocognitive deficits, behavioral disturbances, and cardiometabolic derangements [2]. The most significant risk factor for moderate sleep-disordered breathing in children and adolescents (2-18 years of age) has been associated with obesity [3]. OSAS studies are therefore becoming increasingly more relevant with the epidemic of childhood obesity.

Adeno-tonsillar hypertrophy has been recognized as an important anatomic cause of restriction of the upper airway and contributing to the development of OSAS in obese children. Adeno-tonsillectomy is therefore recommended as the first line of treatment for obese children with OSAS [4-7]. Organ segmentation, especially of tonsils, is needed in analyzing size changes due to surgery and studying treatment effects [8]. Modified and resected organs, such as tonsils after tonsillectomy, are generally a challenge for most of the current multiple organ segmentation methods [9-12].

The Automatic Anatomy Recognition (AAR) methodology based on fuzzy object models (FOMs) we previously reported [13] demonstrated its ability to recognize and delineate multitudes of objects body-wide in CT and MR imagery. We previously tested this methodology on image cases with some pathology where the organs were not distorted significantly. It may be a challenge for fuzzy object models and AAR recognition to deal with the data where organs (tonsils) are distorted or absent due to adeno-tonsillar hypertrophy or adeno-tonsillectomy, since the fuzzy object models are built from normal subjects who commonly have normal tonsils and both tonsils. So it will be useful to study what will happen on those cases with AAR recognition and what strategy to be adopted within AAR to correctly predict absent tonsils. Further, if tonsil removal is correctly predicted, it's not necessary to make delineation on tonsils and AAR

can stop without further proceeding to delineation. In this paper, we present an advancement of AAR to handle organs which may have been modified or resected by surgical intervention (tonsillectomy). Here we focus on MRI of the neck in pediatric OSAS. To our knowledge, no methods have been reported to date for handling significantly deformed or missing organs, especially in neck MRI.

2. METHODS

The proposed method combines AAR-recognition step and support vector machine (SVM) techniques to detect the absence of tonsils in post-tonsillectomy images. The AAR approach consists of the following steps: data acquisition, object delineation on training data sets, fuzzy anatomy model building, recognition by using the models, and delineation by using the pose-adjusted models found at recognition. Compared with the segmentation methods based on multi-atlas and label fusion techniques [17], AAR method allows encoding and further exploiting whatever relationships that exist among objects without imposing any restrictions such as linearity or smoothness. We break up the segmentation task for a given image I into recognition and delineation. The goal of recognition is to determine the object's location and model's adjusted size in the image and the goal of delineation is to determine the object's spatial extent in the image and to mark the precise spatial occupation of the objects in I . SVM is a commonly used classifier which can be linear or non-linear depending on its kernel function. It can effectively deal with binary classification and multiple classification, especially on small size training data sets [14].

The method proposed here mainly includes three steps: building hierarchical models as per AAR principles, object recognition/localization by the AAR methodology, and SVM detection test for objects.

AAR model building

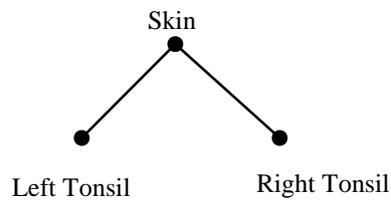


Figure 1. Object hierarchy.

In the AAR approach, given input images I_1, \dots, I_N , for a body region B , in our case neck, for N subjects, and segmentations of objects O_1, \dots, O_L of B in I_1, \dots, I_N , the fuzzy anatomy model of B , $FAM(B)$, is defined to be $FAM(B) = (H, M, \rho, \lambda, \eta)$. Here, H is a hierarchy of organs in B which can be very detailed as described in [13] or simple as illustrated in Figure 1 for our application of the detection of the tonsils. $M = \{FM_\ell : 1 \leq \ell \leq L\}$ is a set of fuzzy models; each fuzzy model FM_ℓ represents the model of object O_ℓ as a fuzzy set estimated over N samples of O_ℓ . ρ denotes the parent-to-offspring relationship in H over N subjects. λ is a set of scale factor ranges $\lambda = \{\lambda_\ell : 1 \leq \ell \leq L\}$, λ_ℓ being a scale factor range for organ O_ℓ over N subjects. η is a host of measurements pertaining to organs in B and the associated images which are used in various processes including recognition and delineation of objects as well as for describing quantitatively the anatomy and image information pertaining to the anatomy of B . $FAM(B)$ handles uncertainties and variability in anatomy naturally and can be integrated with the fuzzy connectedness and other graph based object delineation frameworks naturally. In this paper, the body region neck is defined in axial extent with the superior boundary 6.6 mm above the superior aspect of the globes and the inferior boundary is defined by a level 6.6 mm inferior to the inferior aspect of the mandible.

AAR recognition

Given $FAM(B)$ and a test image I , the recognition method determines the optimal pose (location, orientation, and scale factor) for the object's model for each object in image I so the model falls in place as closely as possible to the actual organ manifestation in I .

To determine the optimum pose for object O_k , a subset of the search space is determined based on variations observed in ρ_k (that is, variation in the relationship of O_k with its parent) and λ_k . This subspace is then sampled at regular discrete intervals, and at each sample pose, a "recognition score" is valuated to examine the degree of match of FM_k with the evidence available in I for O_k . The pose returning the best score is considered to be the pose recognizing object O_k . For the "optimum threshold" method, which is the best among the recognition strategies we have studied [13], given a

specific object O_k , we determine an optimum threshold interval Th_k by rehearsing a recognition trial on the training images that yields the best recognition accuracy. Then with that threshold, we segment the test image I into a rough segmented binary image $J_k(I)$. The optimal pose for O_k then is the pose with which we can achieve minimum difference between the transformed model FM_k^p at pose p and the thresholded image $J_k(I)$.

$$p^* \in \arg \min_p (|FM_k^p(O_k) - J_k(I)| + |J_k(I) - FM^p(O_k)|) \quad (1)$$

Here, image subtraction is done in the sense of fuzzy logic, and $|x|$ denotes the fuzzy cardinality of x , meaning that it represents the sum total of the membership values in x . In searching for the best pose p , we take the model itself as reference for analyzing false positive and false negative portions as in (1).

For the optimal threshold method adopted in AAR recognition, two issues related to image intensity should be considered. One is intensity standardization and another is non-uniformity correction. MR images usually suffer from the problem of intensity non-standardness which implies the lack of a tissue specific numeric intensity meaning of MR images even within the same MRI protocol, for the same body region, for images obtained on the same scanner, and for the same patient. Non-uniformity of image intensity is caused by the imperfections of the imaging device which results in the same tissue appearing very bright and dark in different parts of the image. For all images in our experiments, we first apply non-uniformity correction and then automatic intensity standardization [15, 16].

SVM detection

We use LIBSVM software with a Radial Basis Function kernel and initial parameters set to default values [14]. The output of the SVM classifier indicates whether the tonsil has been found to be absent (value 1) or if it is present in the image (value 0). The feature vectors for SVM detection consist of two components corresponding to the left and right tonsils. Each component indicates the size (fuzzy volume) of the intersection of the fuzzy model with the optimally thresholded image where the pose of the fuzzy model is the optimal found at AAR recognition. SVM training is performed at the model building stage on the training data sets. The idea is that, at an optimal threshold and at optimal recognition, the above fuzzy volume of intersection will be much lower if a tonsil is actually absent than if it is left intact. The proper decision function to employ is determined by the training samples by SVM.

After SVM detection, if tonsils are found to be present, the AAR approach will proceed to delineation of the target object in the given image I . Delineation of objects is performed on I in the hierarchical order as in recognition. The root object, such as skin, is first delineated and then delineation proceeds in the breadth-first order to other off-spring objects until all the objects are delineated. AAR delineation is based on Iterative relative fuzzy connectedness (IRFC) with affinity calculation taking into account object homogeneity, intensity feature, and the model membership at recognition [13].

3. RESULTS

Institutional Review Board approval was obtained for this prospective study. The following three data sets are employed in our experiments: DS1: images of 15 OSAS subjects with all tonsils intact; DS2: image data simulating tonsil removal performed on images of DS1; DS3: real pre- and post-tonsillectomy data sets from 11 subjects. All tonsillectomy subjects are female with age in the range 8-17 yrs.

Simulated tonsillectomy data set DS2 consists of the original 15 unmodified data sets, plus additional 15 data sets for each of the three cases of tonsil removal – left only, right only, and both, yielding a total of 60 image data sets. Tonsil removal is simulated by replacing the standardized MR image intensity values in the region of the tonsil to be removed by intensities which follow a Gaussian distribution whose mean and standard deviation are estimated from the region in the vicinity of the tonsil in that image.

The MR images are all of size 400 x 400 x 35-50 with a voxel size of 0.5 x 0.5 x 3.3 mm³ and axial T2 weighted from a Philips Achieva 3T scanner with TR/TE=8274.3/82.6 msec for DS1 and TR/TE=11929/100 msec for DS3. Fuzzy object models were created from DS1. Leave-one-out cross evaluation is performed for SVM detection on each data set group.

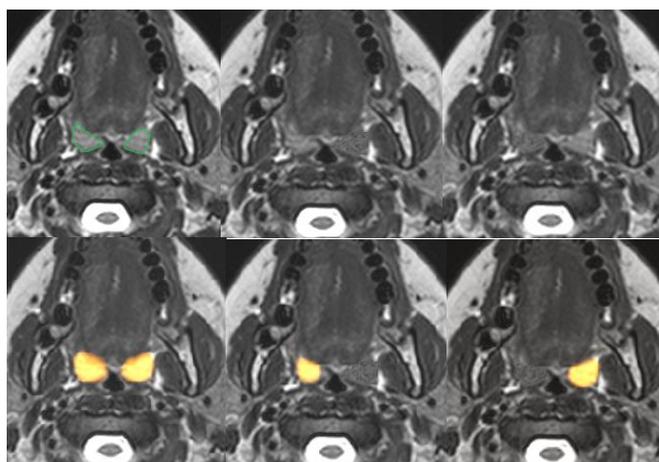


Figure 2. Recognition results from a simulated data set in DS2. Top row: A slice of a simulated tonsillectomy data set before surgery (left), and after removing right tonsil (middle) and left tonsil (right). The boundaries of the tonsils are marked in the first image. Bottom row: Recognition results where the cross section of the fuzzy model is overlaid in yellow on the image. Left to right: no removal, left tonsil removed, right tonsil removed.

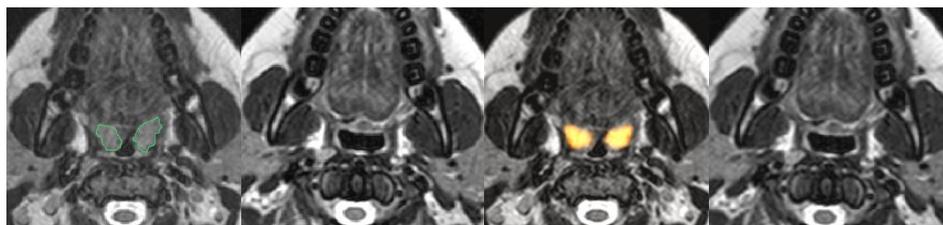


Figure 3. Recognition results from a real patient data set in DS3. Left to right: pre-operative image with the boundaries of the tonsils marked, post-operative image with tonsils removed, recognition result overlaid (in yellow) on pre-op and post-op images (no tonsils detected).

AAR recognition results from simulated and real tonsillectomy data are displayed in Figures 2 and 3 where tonsils are fully removed. Upper airway cross sectional area seems to have increased in the real post-operative image. More quantitative results on the effects of tonsillectomy on OSAS can be found in previous work [8], where the segmentation was done manually.

	Neck Skin	Right Tonsil	Left Tonsil	Mean
Location Error (mm)	3	2.9	2.3	2.73
Standard deviation (mm)	1.2	2.2	2.1	1.83
Size error	1	0.92	0.90	0.94
Standard deviation	0.01	0.11	0.12	0.08

Quantitative results of AAR-Recognition and SVM detection are presented in Tables 1-3. Table 1 shows results for data set DS1 where location error denotes the distance between the centers of ground truth objects and the fuzzy model at optimal recognition. Scale error denotes the ratio of the estimated size to true size. Perfect recognition results would yield location and size errors of 0 and 1, respectively. Since the voxel size is $0.5 \times 0.5 \times 3.3 \text{ mm}^3$, the location error

expressed in voxel units is within one voxel. SVM detection performance on data sets DS2 and DS3 is shown in Tables 2 and 3. The tables also list the results for each sub-group. For both data sets, True positive (TP) rate is around 95% and false positive (FP) rate is close to zero. An interesting observation for one special case in real tonsillectomy data is shown in Figure 4, where tonsils are largely removed but AAR still seems to be able to catch the small remains of tonsils as shown in Figure 4.

Simulated tonsillectomy data set	FN	FP
No removal	0/15	0
Left removed	1/15	0
Right removed	1/15	0
Both removed	0/15	0
All	2/60	0

Tonsillectomy data set	FN	FP
Pre-op	0	0
Post-op	1/11	0
Pre & Post	1/22	0

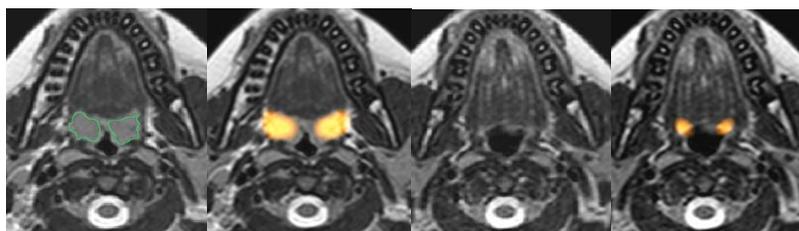


Figure 4. Recognition results from one real patient data set in DS3. Left to right: pre-operative image with the boundaries of the tonsils marked, and the recognition result, post-operative image with partial tonsils removed and the recognition result (with residual tonsils detected).

4. CONCLUSIONS

The previous AAR methodology [13] is extended by combining hierarchical modeling and recognition strategies with machine learning techniques to handle post-surgery situations where objects have been resected or modified surgically. The ability of the AAR approach to detect missing organs is demonstrated on simulated and real post-tonsillectomy MRI image data of the neck. Experimental results show that AAR-Recognition techniques can be combined with machine learning strategies within the AAR recognition framework for good performance in recognizing missing organs, in our case tonsils. The previous recognition performance is maintained achieving an organ localization accuracy of within 1 voxel when the organ is actually not removed. Since new additions are within the computationally very efficient AAR framework, there is no significant additional computational burden.

One limitation of our study is the small number of data sets employed in experiments. However, the data provided preliminary results for a novel way of analyzing tonsillectomy image data in OSAS within the AAR framework. Image quality is improved via automatic non-uniformity correction and intensity standardization approaches in this paper, which may be further improved with the novel interactive non-uniformity correction and intensity standardization approached as shown in [19] and then benefit the proposed predicting approach, perhaps even more.

Acknowledgements

The research reported here is supported by a DHHS grant HL105212.

REFERENCES

- [1] Ogden C.L., Carroll M.D., Kit B.K. and Flegal K.M., "Prevalence of childhood and adult obesity in the United States, 2011-2012," *Journal of the American Medical Association*, 311(8):806-814 (2012).
- [2] Muzumdar H., Arens R., "Physiological effects of obstructive sleep apnea syndrome in childhood," *Respir Physiol Neurobiol*, vol. 188, 370-382 (2013).
- [3] Davidson Ward S. L., Amin R., Arens R., Chen Z.P., Gutmark E., et al., "Pediatric sleep-related breathing disorders: advances in imaging and computational modeling," *IEEE Pulse*, vol. 5, 33-39 (2014).

- [4] Arens R., Sin S., Nandalike K., et al. Upper airway structure and body fat composition in obese children with obstructive sleep apnea syndrome. *Am J Respir Crit Care Med.* 1vol.183, 782–787 (2011).
- [5] Dayyat E., Kheirandish-Gozal L., Maarafeya M.M., Gozal D., “Obstructive sleep apnea in children: relative contributions of body mass index and adenotonsillar hypertrophy,” *Chest*, vol. 136, 137–144 (2009).
- [6] Gordon JE, Hughes MS, Shepherd K, et al., “Obstructive sleep apnoea syndrome in morbidly obese children with tibia vara,” *J Bone Joint Surg*, vol. 88, 100-103 (2006).
- [7] Mitchell RB, Kelly J. “Adenotonsillectomy for obstructive sleep apnea in obese children,” *Otolaryngol Head Neck Surg*, vol. 131, 104–108 (2004).
- [8] Nandalike K., Arens R., et al., “Adeno-tonsillectomy in Obese Children with Obstructive Sleep Apnea syndrome: Magnetic Resonance Imaging findings and considerations,” *Sleep*, 36(6):1-7 (2013).
- [9] Suzuki M., et al., “Analyses of Missing organs in Abdominal Multi-organ segmentation,” *Abdominal Imaging*, vol. 7029, 256-263 (2012).
- [10] Suzuki M., et al., “Multi-organ segmentation with missing organs in Abdominal CT images,” *MICCAI*, vol. 7512, 418-425 (2012).
- [11] Zhou Y., Bai J., et al., “Multiple abdominal organ segmentation: an atlas-based fuzzy connectedness approach,” *IEEE Trans. Info. Technol. BioMed.*, 11(3):348-352 (2007).
- [12] Lee HP et al., “Simulated-based joint estimation of body deformation and elasticity parameters for medical image analysis,” *IEEE Trans. Medical Imaging*, 31(11):2156-2168 (2012).
- [13] Udupa J.K., et al., “Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images,” *Med Image Anal*, 18:752-771 (2014).
- [14] Chang C-C., Lin C.J. “LIBSVM: A library for support vector machines,” *ACM Trans. IST*, 2(27):1-27 (2011).
- [15] Nyúl LG, Udupa J.K., “On standardizing the MR image intensity scale,” *Magn. Reson. Med.*, 42 (6):1072-81 (1999).
- [16] Ying Z., Udupa J.K., et al., “Image background inhomogeneity correction in MRI via intensity standardization,” *Comp. Med. Imag. and Graph.*, 33(1):7-16 (2009).
- [17] Bai W., Shi W.Z., Ledig C., Rueckert D., “Multi-atlas segmentation with augmented features for cardiac MR images,” *Medical Image Analysis*, vo. 19, 98-109 (2015).
- [18] Wang H., Suh W., Yushkevich P., “Multi-Atlas Segmentation with Joint Label Fusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):611-623 (2013).
- [19] Tong Y.B., Udupa J.K., et al., “Interactive Non-Uniformity Correction and Intensity Standardization of MR Images,” *Proceedings of SPIE, Medical Imaging 2015*, vol. 9414 (to appear).