

Automatic Cephalometric Evaluation of Patients Suffering from Sleep-Disordered Breathing

Lior Wolf¹, Tamir Yedidya¹, Roy Ganor¹, Michael Chertok²,
Ariela Nachmani³, and Yehuda Finkelstein^{3,4}

¹ School of Computer Science, Tel-Aviv University

² School of Engineering, Bar-Ilan University

³ Center of Cleft Palate and Craniofacial Anomalies, Meir Medical Center

⁴ Palate Surgery Unit, the Department of Otolaryngology: Head and Neck Surgery, Meir Medical Center, affiliated with Sackler School of Medicine Tel-Aviv University

Abstract. We address the problem of automatically analyzing lateral cephalometric images as a diagnostic tool for patients suffering from Sleep Disordered Breathing (SDB). First, multiple landmarks and anatomical structures that were previously associated with SDB are localized. Then statistical regression is applied in order to estimate the Respiratory Disturbance Index (RDI), which is the standard measure for the severity of obstructive sleep apnea. The landmark localization employs a new registration method that is based on Local Affine Frames (LAF). Multiple LAFs are sampled per image based on random selection of triplets of keypoints, and are used to register the input image to the training images. The landmarks are then projected from the training images to the query image. Following a refinement step, the tongue, velum and pharyngeal wall are localized. We collected a dataset of 70 images and compare the accuracy of the anatomical landmarks with recent publications, showing preferable performance in localizing most of the anatomical points. Furthermore, we are able to show that the location of the anatomical landmarks and structures predicts the severity of the disorder, obtaining an error of less than 7.5 RDI units for 44% of the patients.

1 Introduction

Sleep-Disordered Breathing (SDB) consists of a continuous upper airway resistance with clinical indications ranging from snoring to Obstructive Sleep Apnea (OSA). In this disorder, the breathing temporarily stops during sleep as the throat muscles relax and block the patient's airway. The patient then wakes up in order to restart his breathing, and the quality of sleep is impaired. OSA symptoms include daytime sleepiness and fatigue, as well as an increased risk of cardiovascular disease, stroke, high blood pressure, arrhythmias, and diabetes.

The prevalence of sleep apnea is estimated to be 3.3% in the adult male population and increasing to almost 5% in the middle age group [1]. The objective diagnosis of OSA and its severity, and establishing a baseline for future changes require a referral to a sleep laboratory. The evaluation involves an overnight stay and monitoring of the breathing during sleep following an analysis by a specialist. While the resulting report is considered the gold standard in current diagnosis, its reliability is questionable since

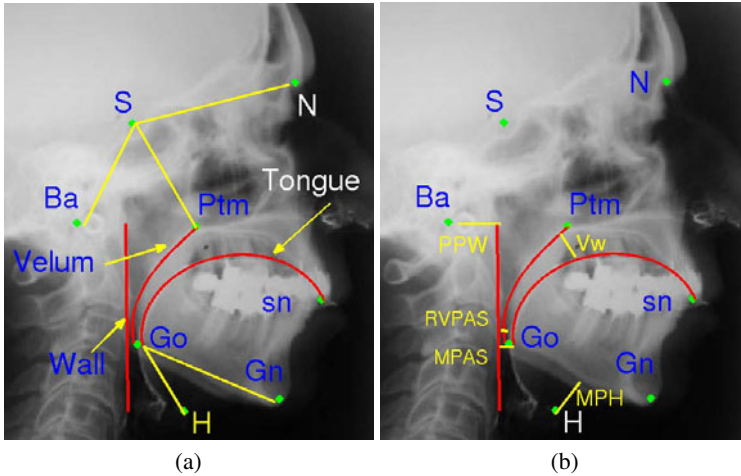


Fig. 1. Landmarks and anatomical structures. The 8 landmarks are marked with green points. (a) The tongue, velum and Pharyngeal Wall are marked in red (from right to left). The three angular measurements (in yellow) are: bony nasopharyngeal angle (**BaSPtm**), skull base angle (**BaSN**) and mandibular plane hyoid angle (**GnGoH**). (b) The five line measurements marked in yellow. **Vw** is the maximal width of the velum (velar width). **RVPAS** is the minimum distance between the velum and the posterior pharyngeal wall. **MPH** is the distance between landmark H (hyoid) to line GnGo (mandibular plain). **PPW** is the distance from the basion (Ba) to the wall (defined slightly differently from [4]). **MPAS** is the minimum distance between the tongue and the posterior pharyngeal wall. Notice that landmark sn (end of tongue) is different from Prn, which is detected in some previous work.

patients may fail to sleep naturally at the laboratory [2,3]. Moreover, the main drawback of the polysomnography is that it is only a descriptive tool which does not provide any information regarding the pathophysiology of SDB. Diagnosis of the anatomic factors that predispose the airway to collapse during sleep is required for tailoring the best treatment modality to the individual patient. Therefore, due to the high prevalence of OSA, its implications of the public health, the cost, length, and inherent limitations of the sleep-laboratory examination, an imaging based diagnosis might be beneficial.

Finkelstein et al. [4] lay the groundwork for 3D cephalometric analysis of SDB by the instillation of contrast media and adding a frontal view to the traditionally used lateral radiographs, demonstrating that worsening of SDB was generally associated with increased numbers of discovered *compromised cephalometric parameters*. The cephalometric analysis is performed by manually marking the anatomical landmarks and measuring various angles and distances, and interpreted by a skilled expert.

Here, we (i) automate the process described in [4]: The cephalometric landmarks are detected automatically using a new registration method, followed by an automatic detection of the anatomical structures; and (ii) the Respiratory Disturbance Index (RDI) that measures the severity of the SDB is predicted based on the image measurements. Therefore we switch from manual marking that requires a high level of expertise to an automatic one, and at the same time provide predictability and not just correlation.

Previous Work. To our knowledge, there is no existing research on automatic analysis of cephalometric images for SDB. Previous research focused on cephalometric landmark detection for orthodontics. The more recent methods use Active Shape Models (ASM) or Active Appearance Models (AAM) [5,6,7,8,9]. Typically, models with a large number (~ 250) of landmarks are learned, an initial localization is provided by heuristically detecting the easiest landmarks (e.g., Prn and Gn in [9]), and ASM/AAM is then used for fine-tuning.

2 Triplet-Based LAF for Image Registration

We represent each image by a set of local regions, each associated with its own affine coordinate frame. Such a representation is sometimes called a Local Affine Frames (LAF) representation, and was introduced in [10]. In the literature LAFs are derived based on Maximally Stable Extremal Regions (MSERs) [11], which do not provide desirable output on cephalometric images. Here, triplets of keypoints are used as the basis for our LAFs, each detected using the DoG interest point operator [12].

1. Keypoint detection. The SIFT DoG operator [12] is applied to the image obtaining a large group of keypoints p_1, \dots, p_n (n varies between images).

2. Random triplet selection. A large number of triplets ($n_T = 10,000$ in all of our experiments) is selected by random.

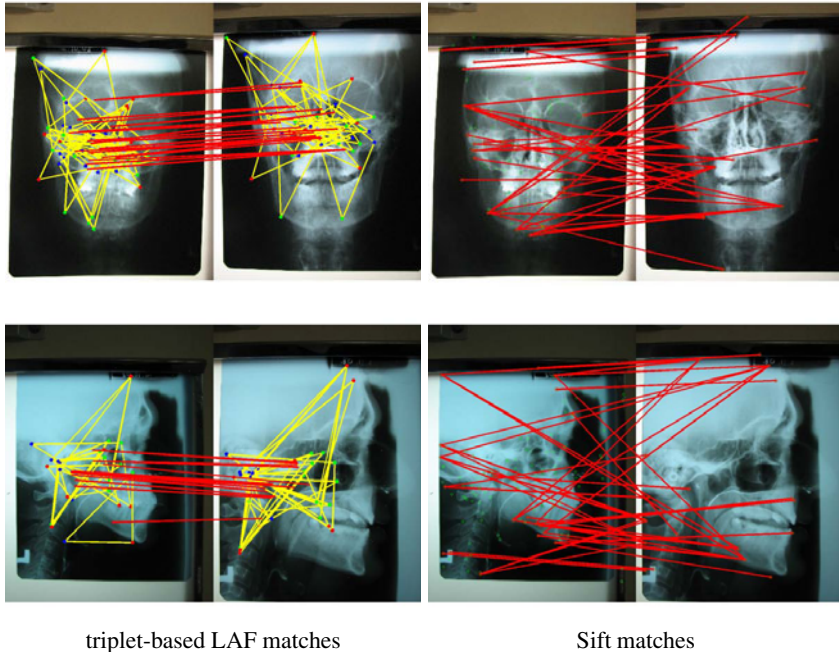


Fig. 2. Top triplets-based LAF matches (left) and SIFT matches (right). Coherent matches produce parallel lines. As can be seen the proposed LAF method provides matches that are much more coherent. The top row demonstrates results on frontal cephalometric images.

3. Triplet ordering. The points of each triplet are ordered to provide a 6 fold reduction in the number of potential triplets. The order is obtained by computing for each keypoint the local SIFT descriptor, and then projecting the descriptor vector to the first principle component obtained from a large group of SIFT features that were collected beforehand. Each triplet $T_i = \{p_{i,1}, p_{i,2}, p_{i,3}\}, i = 1..n_T$ is then ordered such that $p_{i,1} < p_{i,2} < p_{i,3}$, where the order is given by the projection of the SIFT vector.

4. Region of interest extraction. We compute an affine transformation that maps the three points of each triangle to the coordinates $(1, 1), (128, 1), (64, 128)$, and use it to warp the image. Then an enclosing square of corner coordinates $(1, 1)$ and $(128, 128)$ is used to obtain local appearance information for each triplet.

5. Appearance encoding. We record for each triangle the histogram of edge directions at the enclosing square, by using the SIFT descriptor [12].

After each image (I) is encoded as a collection of triplets and matching ROI descriptors, it can be compared to another image (J). The matching score between pairs of triplets, one from I and one from J , is provided by the distance measure of the SIFT descriptors. The highest scoring pairs provide information on the overall match quality. Figure 2 compares triplet-based matches to matches based on DoG (SIFT) keypoints.

3 Landmarks Detection

We are given m_t training images that are manually marked with the relevant anatomical landmarks. Two different sets of tracings and measurements were conducted by A.N. and Y.F. for each roentgenogram. The mean values of the two sets of measurements were used. Following [4], we employ a group of 8 cephalometric landmarks: $\mathbf{m} = \{Gn, Go, Ba, S, N, sn, Ptm, H\}$ (see Figure 1). Given an unannotated test image I_t , we employ the registration technique of Section 2 in order to find a multitude of matching LAFs between I_t and each of the training images $I_j, j = \{1, 2, \dots, m_t\}$. Each LAF matching provides: (i) a score computed from the distance of the SIFT descriptors; (ii) a training image index, and (iii) an affine transformation between that training image and the test image. We first select the 50 LAFs with the highest scores from among the pool of all LAFs. Then, the associated affine transformations are used to transform the locations of the landmarks from the training images to the test image.

This process results in a collection of 50 projected locations per anatomical point in the test image (see Fig. 3(a)). A single location per landmark is obtained by employing a mean shift algorithm [13] to locate the mode of each group of landmark projections.

Next, a fine-tuning step is applied in order to find the exact location of each landmark. For some landmarks, such as Go, localization can be improved by seeking the nearest edge points. For others, such as H and Ba that are not necessarily located on visible image edges or defined by a specific texture, a more sophisticated approach is needed. We choose to perform the fine-tuning step to all landmarks using the same local appearance based method.

To that end, we create a set of templates for each anatomical landmark, by projecting for each LAF the associated training image to the test image, and cropping a template around each anatomical point. For each landmark, an SVM classifier is trained to distinguish the relevant templates from nearby templates. The resulting detector is applied in

the vicinity of the initial mean-shift based estimation, and the maximum detector score is used to select the final location.

Once the landmarks are localized, we can evaluate the three angular measurements of [4]: the skull base angle (BaSN); the bony nasopharyngeal angle (BaSPtm); and the mandibular plane hyoid angle (GnGoH) (see Fig. 1(a)).

4 Detection of Cephalometric Structures

Next, we model the tongue, the velum and the posterior pharyngeal wall. These structures are critical for the evaluation of SDB since they are involved in the narrowing of airspace and other anatomical changes. For each structure, we fit a suitable geometric shape, whose parameters are searched in a range dictated by the training images.

In order to reduce the variability in the parameters of the structures, the images are first aligned with accordance to the locations of the eight anatomical feature points. For each landmark, we find the mean location in all training images $\bar{\mathbf{m}} = \sum_{i=1}^{m_t} \mathbf{m}_i$. For a given image I_j (training or testing), we compute the affine transformation that minimizes the least-squares error from \mathbf{m}_j to $\bar{\mathbf{m}}$ and apply it on I_j and its landmarks. All anatomical structures are detected using the aligned images, however, the actual measurements are performed after aligning the detections back to the original coordinate system.

Tongue. The relevant part of the tongue is from its intersection with the contour of the chin towards the mouth's end (sn). We found out that an ellipse $E = ax^2 + bxy + cy^2 + dx + ey + f = 0$ provides a good fitting in most cases, however, simpler models, such as a circle do not fit well. Since an ellipse is defined by five degrees of freedom, and since the two landmarks Go and sn are known to be on the tongue, three more degrees of freedom are to be evaluated.

One possible solution would be to find edge points in the vicinity of the tongue and use RANSAC to fit an ellipse, however, the velum is similar to the tongue and it is easy to confuse between the two. Instead, we perform a search in a parameter space that is defined by the anatomical points.

Observe Figure 3(b). We draw three lines: one from Gn to Ba, one from Gn to Ptm and a third line forming an angle twice as large as the angle BaGnPtm. The ellipse is defined by the points Go and sn and the intersection of its top part with the the three lines. The distances of these three intersection points from Gn are used as the ellipse parameters. The suitable range for each parameter is estimated from the training images. This set of three parameters has a much more compact range than the range of the generic ellipse parameters, since the tongue might be tilted at various angles depending on the anatomy of each patient. The actual fitting is performed by an exhaustive search in these ranges, and the quality of each hypothesis ellipse is measured by the edge intensity along its path.

Velum and Pharyngeal Wall. In order to measure the Velar Width (VW) as in [4] (see Figure 3(c)), we detect the velum in the region to the left of landmark Ptm and find the maximum distance to the tongue. No simple geometric shape fits well to the velum, since it may curve strongly towards its top part. We model the velum using a cubic B-spline with four knots [14]. We set the coordinates of the first and last knots at Ptm and

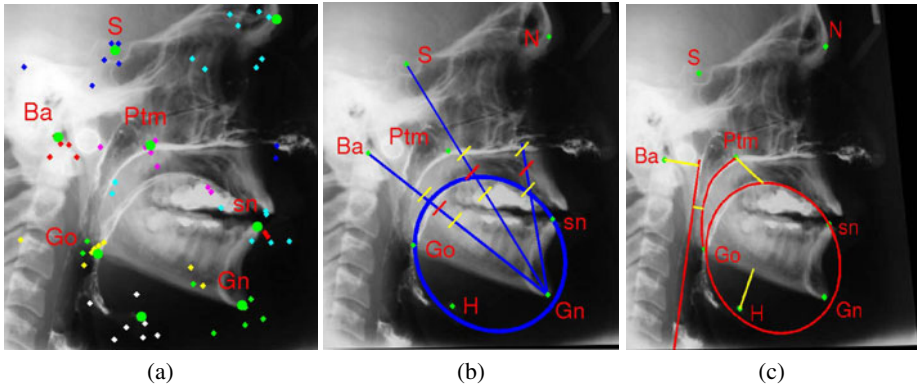


Fig. 3. The process of anatomical structures fitting. (a) Projected locations - the chosen points per each anatomical landmark are marked in the same color and the green dot is the correct location. One can see that at least one location per landmark is accurate. (b) Tongue Fitting. The tongue is modeled by using an ellipse. The green dots are the automatically detected landmarks from section 3. The lines that are used to define the three ellipse parameter points are plotted. The yellow marks on each line bound the valid searching distance from Gn and the red marks are the average distance from Gn. (c) Fitting of all structures. The yellow lines correspond to Vw, MPH, RVPAS and PPW. The tongue is detected correctly even though the teeth are not touching.

Go respectively. Using similar ideas to the tongue fitting, we draw two lines from the Gnathion (Gn) that intersect the velum. Then we measure the valid range from the tongue to the velum in the training images. These ranges define the searching area for the two additional knots and we iteratively fit B-splines to find the best fitting. The pharyngeal wall is modeled by a straight line, since it provides an accurate enough estimation in the region relevant to the line measurements. The model is parameterized by the intersection of the pharyngeal wall line with the two lines GnBa and GnGo, and the search is performed as above in a range dictated by the training images. Fig. 3(c) shows the detection of the three structures (in red) and the yellow lines refer to the 4 line measurements used.

5 Results

We have collected 70 cephalometric images of patients with varying degrees of OSA. All images were taken in accordance with recognized standards: the head was oriented in the Frankfurt horizontal plane and stabilized with a head-holder; the teeth were in habitual occlusion with lips together, and the tongue relaxed in the floor of mouth. Exposures were taken during mid-expiration following a moderately deep breath. Therefore, the variation in pose is minimal. The database is challenging since the quality varies, and in some cases the contrast is poor and the visibility of the anatomical structures is limited.

We randomly split the dataset to 25 testing images and 45 training images and repeat the experiments 10 times. The results discussed below are mean results that were computed automatically on the testing images. First, we compare the accuracy of the landmarks discussed in [4] with the manual detection and three other approaches (see Table 1). As far as we know, the datasets used in previous work are not available. Therefore,

Table 1. Automatic landmarks detection results. In parenthesis, when available: ^a Rueda et al [8], ^b Yue et al [9], and ^c Hutton et al [6]. Notice the high accuracy of S, Ptm, N and Gn. Landmark H has a relatively low detection rate. However, large variation exists also between clinicians marking this landmark. (In [9]the same landmarks are named differently, thus we compare Gn with Me, Ptm with Pns and Ba with Co).

Landmark	mean (mm)	≤ 2mm(%)	≤ 3mm(%)	≤ 4mm(%)	≤ 5mm(%)
S	0.48(2.29 ^a ,5.5 ^c)	100(39 ^a ,76 ^b)	100(70 ^a)	100(78 ^b)	100(98 ^a)
Ptm	0.53(2.67 ^a ,5.0 ^c)	100(37 ^a ,83 ^b)	100(69 ^a)	100(92 ^b)	100(92 ^a)
N	1.32(5.6 ^c)	91(86 ^b)	96	96(89 ^b)	96
Gn	1.21(1.58 ^a ,2.7 ^c)	83(73 ^a ,98 ^b)	83(86 ^a)	87(100 ^b)	91(99 ^a)
Go	2.32(3.88 ^a ,5.8 ^c)	52(26 ^a ,86 ^b)	74(44 ^a)	83(94 ^b)	95(67 ^a)
Ba	2.06(2.7 ^a)	65(38 ^a ,69 ^b)	74(68 ^a)	78(76 ^b)	87(92 ^a)
H	6.4	13	22	30	39

the comparison provides merely a general impression. We have achieved very accurate results for landmarks S, Ptm, Gn and N, even though S and Ptm are not located on clear edges. Comparing to [8] and [6], we have achieved better mean error for all reported landmarks. Comparing with [9], we have better results for S and Ptm and similar results for N, Ba and Gn. Regarding landmark Go, the authors of [9] assumed that the landmark can be found by tracing the edge that starts from Gn (Me in their paper), however, this assumption does not hold in many of our images, where the two sides of the mandible are visible (see the double line in Fig. 1). In such cases, the location of Go is not situated on this edge point. Results for H are not reported in previous work.

We also compare our automatic line measurements with the manual ones. The mean absolute errors and standard derivations ($x \pm y$) are: Vw 0.94 ± 0.78 , RVPAS 1.3 ± 1.3 , PPW 2.3 ± 2.2 and MPH 3.8 ± 3.4 . The errors in computing Vw and RVPAS are very low. The error in detecting MPH is due to the difficulty of finding landmark H. However, considering the inter-observer variance between clinicians in marking this landmark, our results are in the acceptable range.

Next, we predict the RDI and hence the severity of the SDB. In our testing images, the RDI varies from 5 (borderline healthy) to 77 (severe OSA). The prediction is performed via a linear Support Vector Regression model, that is based on either: (1) The three angles measured; (2) The 4 line measurements; (3) a concatenation of (1) and (2);

Table 2. RDI prediction results. See detailed description of the methods in section 5. Column 2: Mean squared RDI error; Columns 3 and 4: the percentage of cases that have a relatively low RDI error (≤ 7.5) and those with a more moderate error (≤ 15). The results could be compared to the night-to-night variability in RDI computations at sleep clinics which displays a standard deviation of 7.2 [2].

Method	Mean Squared Error (MSE)	error ≤ 7.5(%)	error ≤ 15(%)
Angles	324	26	61
Lines	393	30	53
Angles+Lines	361	44	57
BMI	511	17	43

and as a baseline, (4) The Body Mass Index (BMI). Table 2 summarizes the regression results. It is well known that BMI and RDI are well correlated. Nevertheless, the cephalometric measurements are much more predictive of the patients RDI. We therefore corroborate the claim of [4] that the anatomical structures are informative for OSA diagnosis and support the link between available airspace and the degree of the disorder.

6 Conclusions

Previously it was shown that manual measurements of cephalometric images are correlated with the Respiratory Disturbance Index. In this paper, we show, for the first time, that RDI measurements can be predicted from cephalometric images, moreover, our measurements are obtained automatically. Therefore, we make a significant advancement toward an imaging based OSA diagnosis tool. The detection of the underlying landmarks is based on a new registration technique that is potentially useful for several other structure detection applications.

References

1. Young, T., Peppard, P.E., Gottlieb, D.J.: Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective. *Am. J. Respir. Crit. Care Med.* 165(9), 1217–1239 (2002)
2. Mosko, S., Dickel, M., Ashurst, J.: Night-to-night variability in sleep apnea and sleep-related periodic leg movements in the elderly. *Sleep* 11(4), 340–348 (1988)
3. Ahmadi, N., Shapiro, G.K., Chung, S.A., Shapiro, C.M.: Clinical diagnosis of sleep apnea based on single night of polysomnography vs. two nights of polysomnography. *Sleep and Breathing* 13(3), 221–226 (2009)
4. Finkelstein, Y., Wexler, D., Horowitz, E., Berger, G., Nachmani, A., Shapiro-Feinberg, M., Ophir, D.: Frontal and lateral cephalometry in patients with sleep-disordered breathing. *Laryngoscope* 111, 623–641 (2001)
5. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 681–685 (2001)
6. Hutton, T.J., Cunningham, S., Hammond, P.: An evaluation of active shape models for the automatic identification of cephalometric landmarks. *Eur. J. Orthodont.* 22 (2000)
7. Kafieh, R., Mehri, A., Sadri, S.: Automatic landmark detection in cephalometry using a modified active shape model with sub image matching. In: *ICMV 2007*, pp. 73–78 (2007)
8. Rueda, S., Alcaniz, M.: An approach for the automatic cephalometric landmark detection using mathematical morphology and aam. In: Larsen, R., Nielsen, M., Sporning, J. (eds.) *MICCAI 2006*. LNCS, vol. 4190, pp. 159–166. Springer, Heidelberg (2006)
9. Yue, W., Yin, D., Li, C., Wang, G., Xu, T.: Automated 2-d cephalometric analysis on x-ray images by a model-based approach. *IEEE. Tran. Biomed. Eng.* 53(8) (2006)
10. Obdržálek, S., Matas, J.: Object recognition using local affine frames on distinguished regions. In: *The British Machine Vision Conf.* (2002)
11. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: *The British Machine Vision Conf.*, pp. 384–393 (2002)
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
13. Cheng, Y.: Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 790–799 (1995)
14. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: *Numerical Recipes*, 3rd edn. *The Art of Scientific Computing*. Cambridge University Press, Cambridge (2007)