Nose breathing or mouth breathing? A thermography-based new measurement for sleep monitoring

Zhengjie Huang, Wenjin Wang, Gerard de Haan
Eindhoven University of Technology
Eindhoven, The Netherlands
z.huang2@student.tue.nl, wwang@tue.nl, g.d.haan@tue.nl

Abstract

Nose breathing is preferred during sleep, although health issues may cause a subject to breathe through the mouth, and long-term mouth breathing may raise other health issues like sleep apnea. This paper proposes a first-ever classification of nose breathing and mouth breathing using the thermography of the subject. The measurement uses the relative temperature variations of different facial regions to classify mouth or nose breathing. This measurement is particularly health-/well-being relevant as it can be used as an early sign for sleep disorders or an indicator of sleep quality. An end-to-end processing flowchart has been provided for proof-of-concept validation on real-life recordings of thermal videos. Eight volunteers participated in our experiments and our proposed method achieved an overall classification accuracy of 91% *in ideal lab conditions.*

1. Introduction

Healthy people breathe with both nose and mouth. The nose can warm up and moisturize air from the environment. Also, the chemicals produced by the nose improve oxygen absorption in the lung. Breathing with mouth becomes necessary because of a blocked nose or high-intensity sports. Some people breathe with mouth occasionally while some breathe with mouth almost exclusively which in the long term can lead to a number of health issues like bad breath, periodontal disease, throat and ear infections [4], palatine, and pharyngeal tonsils hypertrophy [11]. It is even worse for children. A study consisting of 661 children participants aged from 6 to 12 years old shows that 26.8% of them are breathing with their mouth [16] and their facial growth can be affected that leads to unattractive facial features [6] if not treated in time. Furthermore, up to 42% of mouth breathers also have apnea according to a study [5]. Therefore, the mouth-or-nose breathing classification is important for the following reasons: early signs of mouth breathing can be captured by overnight monitoring for prevention purposes; the ratio of mouth breathing can be observed for evaluation of recovery from mouth breathing.

2. Methodology

In this section, we will give a detailed description of our processing flowchart as shown in Fig 1.

2.1. ROI extraction

To enable the measurement of the nose and mouth area, we first need to locate the nose and mouth, i.e., extract the ROI. Our ROI extraction is composed of three steps which are face detection, facial landmark localization, nose and mouth extraction.

2.1.1 Face detection

Face detection on RGB images has been advanced, by deep learning techniques and large annotated datasets, to an almost mature status. Nevertheless, due to the substantial differences between thermal images and RGB images, methods that work for RGB images do not necessarily work well for thermal images. Therefore, researchers are still working on robust face detection methods on thermal images. Pereira *et al*. [13] proposed to use the Otsu's multi-level threshold [12] to segment the image into multiple classes. It takes advantage of the fact that the face is usually the warmest object in the image. However, its performance degrades severely in scenes cluttered with objects of different temperatures and it includes some unwanted areas like neck area. The number of levels to segment is also dependent on the actual scene. Furthermore, their method relies on high-resolution thermal cameras that are many times more costly than those low-resolution ones. To obtain a Rectangular region of the face, Filipe *et al*. [3] project the image horizontally and vertically and calculate the maxima and minima of the projections to determine the start and end index of the face area. Marzec *et al*. [10] explored the characteris-

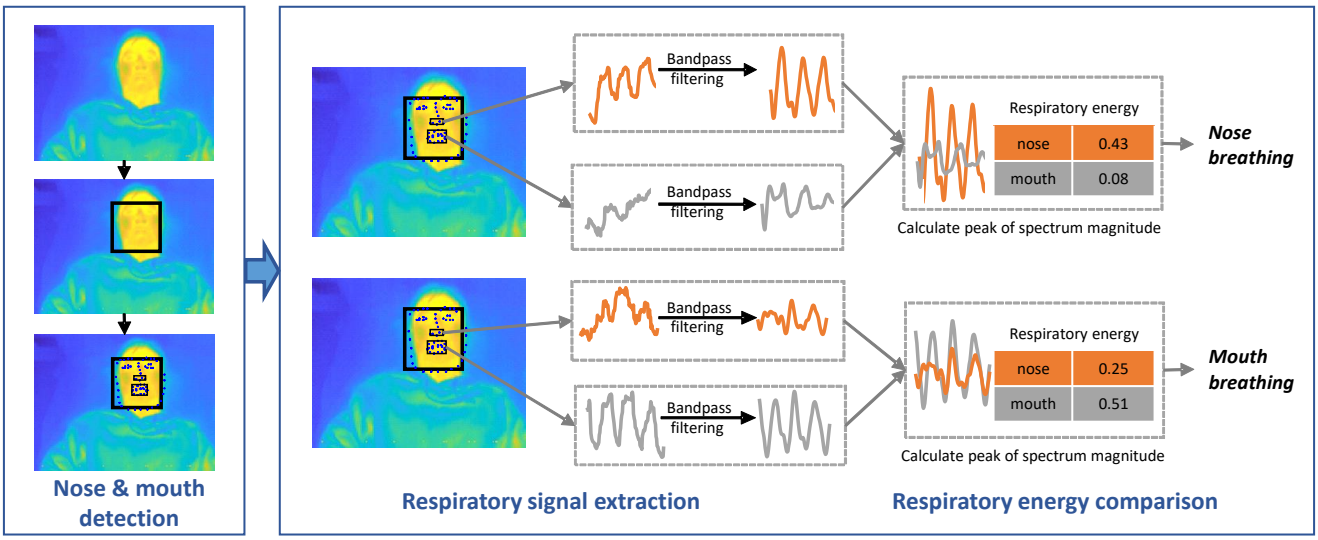


Figure 1: The flowchart for the proposed nose-or-mouth breathing classification consists of two main parts: (i) detect the face and facial landmarks to extract the nose and mouth region of interest; (ii) extract respiratory signals (temperature variations) from nose and mouth areas respectively and compare their respiratory spectra of both regions to arrive at the nose/mouth breathing classification.

tics of thermal facial images. They assumed a few rules based on general temperature distribution on faces and facial anatomy that are applicable when detecting faces and selected points of a face. However, those rules are meant for frontal faces and do not hold for side views. Histograms of Oriented Gradient (HOG) descriptors began to gain popularity from 2005 after its success in human detection [2]. For classification purposes, the Support Vector Machine [1] classifier was also used in that paper. SVM is widely used as a binary or multi-class classifier due to its capability in separating data into clusters. Kopaczka *et al*. conducted experiments on face detection and proved that machine learning-based methods are superior to specialized knowledge-based methods [8]. Therefore, we chose to utilize the maturity offered by machine learning-based techniques that have already been proven to be successful on RGB images.

Since most thermal-based face detection methods use heuristics and assume a clear scene, their performance degrades severely with the presence of clutter and often include some unwanted area like the neck. Those limitations can be overcome with the utilization of facial features along with large training data. Therefore, the face detector we used in this paper is a HOG-SVM detector trained on over 2000 thermal images [7]. To accommodate the face detection for images of different sizes, we scale the image if no face can be detected in its original size.

2.1.2 Facial landmark localization

Facial landmarks can be helpful for tasks involving facial features like face recognition, emotion analysis, etc. In our case, instead of training two detectors for detecting nose and mouth separately, we use the landmarks to locate the ROI at once. Moreover, the facial anatomy can also increase confidence in ROI extraction such that things like the nose is under the mouth will not happen.

Existing facial landmark models can be roughly divided into three categories that are Constrained Local Model (CLM), holistic models, and machine learning-based models [18]. The development of facial landmark localization on RGB images has also been significant after deep learning came into play. Its application on thermal images is still in an early stage though. Moreover, thermal facial images are low in contrast and lack texture information due to the relatively uniform temperature distribution of a face. Thus, existing facial landmark models trained on RGB images do not work on thermal images. Kopaczka *et al*. [7] published a database of fully annotated thermal images, such that it is possible to train machine learning models for different purposes. They also trained an Active Appearance Model (AAM) on thermal images for facial landmark localization and then used the facial features for emotion classification. Furthermore, they evaluated the performance of an AAM which belongs to a holistic model and a Deep Alignment Network (DAN) [9] based model respectively. They showed that DAN outperforms AAM in many aspects including accuracy and speed. DAN also shows promise by

outperforming the other two machine learning-based models - Multi-Task CNN and Patch-based fully convolutional neural network classifier (PBC) in experiments conducted by Poster *et al*. [14].

There is very limited research about facial landmark localization on thermal images to our knowledge. Considering that DAN has been proven feasible [7] and works better than other methods on thermal images [14], we choose DAN as our facial landmark model. As a pre-processing step, we down-scaled the thermal images of a thermal dataset [7] to accommodate our needs for facial landmark localization on low-resolution images. Then we trained the DAN on those down-scaled images and use it as our facial landmark model.

2.2. Mouth-or-nose breathing classification

The area used for air exchange has a higher temperature variation because the air inhaled from the environment is usually colder than the air exhaled through mouth or nose given the condition that the room temperature is stable and lower than the body temperature, which is usually the case. Therefore, the temperature variations caused by air exchange can be used for classification.

2.2.1 Respiratory signal extraction

Since the classification relies on temporal variations, it involves a video sequence. We choose to use a window of 15 seconds as a result of a trade-off between response time and accuracy.

We take the average value of the nose and mouth area from each frame to compose the time-series signal:

$$\begin{cases} \mu_{\mathrm{n},k} = \frac{1}{S(V_{\mathrm{n}})} \sum_{(p,q)\in V_{\mathrm{n}}} i(p,q,k) \\ \mu_{\mathrm{m},k} = \frac{1}{S(V_{\mathrm{m}})} \sum_{(p,q)\in V_{\mathrm{m}}} i(p,q,k) \end{cases} \tag{1}$$

where $i(p,q,k)$ represents the intensity of pixel at position (p,q) of video frame k; V_n and V_m represents of pixel collection of the nose area and mouth area respectively. $S(V)$ represents the number of pixels in V.

In addition, inspired by [17], we also take the standard deviation of the nose and mouth area from each frame for complementary use:

$$\begin{cases} \sigma_{\mathrm{n},k} = \sqrt{\frac{1}{S(V_{\mathrm{n}})-1} \sum_{(p,q)\in V_{\mathrm{n}}} \left| i(p,q,k) - \mu_{n,k} \right|^2} \\ \sigma_{\mathrm{m},k} = \sqrt{\frac{1}{S(V_{\mathrm{m}})-1} \sum_{(p,q)\in V_{\mathrm{m}}} \left| i(p,q,k) - \mu_{m,k} \right|^2} \end{cases} \tag{2}$$

Then, we construct four types of windows: windows of average intensity and intensity standard deviation from both the nose and mouth areas in a 15 seconds time interval. The window slides through the whole video starting from the first frame, generating a signal for further processing.

$$\begin{aligned} R_k &= s_k | s_{k+1} | ... | s_{k+N-1} \\ R_{k+1} &= s_{k+1} | s_{k+2} | ... | s_{k+N} \end{aligned} \tag{3}$$

where s_k refers to $\mu_{\mathrm{m},k}$, $\sigma_{\mathrm{m},k}$, $\sigma_{\mathrm{n},k}$, or $\mu_{\mathrm{n},k}$; R_k is the k-th respiratory window; N refers to the number of frames in a window (450 in our case); Symbol | means signal concatenation.

Furthermore, each window is normalized such that signals within the window have a mean value of 0 and a standard deviation of 1.

$$\hat{s}_i = \frac{s_i - \mu_s}{\sigma_s} \tag{4}$$

where μ_s and σ_s are the average and standard deviation of all signals in a window respectively.

Due to sensor drift and turbulence from the environment, there are inevitably noises that are outside the respiratory rate (RR) range which is 12 to 18 cycles per minute (cpm) [15] in original signals. To exclude these effects, we filter all signals outside the frequency range of possible RR. We also widen the range to 12 to 40 to include some abnormal cases. A second-order Butterworth filter is used such that signals of frequency lower than 0.167 Hz and higher than 0.667 Hz (corresponds to 12 cpm and 40 cpm) are filtered from windowed respiratory signal R_i.

2.2.2 Respiratory energy comparison

After filtering out the noises, the remaining signals consist of mainly breathing signals. By comparing the time-domain signals, specifically, respiratory signals of the nose area and mouth area, we can know which one contributes to air exchange or if both of them do.

As aforementioned, we use a window of 15 seconds (450 frames) and slide it through the whole video with a stride of 1 frame. Therefore, we get a classification result for each frame in the video except for the last 15 seconds. For each window R_i, we get its spectrum Y_i by 1D Fourier Transform:

$$Y_i = \mathcal{F}(R_i) \tag{5}$$

We then choose the highest magnitude E within the respiration band as an energy level indicator.

$$E_i = \max_{f_{min} \le f \le f_{max}} (|Y_i(f)|) \tag{6}$$

where $|Y_i(f)|$ refers to the magnitude spectrum at frequency f; f_{min} and f_{max} refer to the minimum and maximum frequency of respiration band which are 0.167Hz and 0.667Hz respectively. Note that since we have windows obtained from average intensity and standard deviations, the energy indicators can be obtained from those two traces separately.

During nose breathing, the respiratory signal of the nose area will be stronger than that of the mouth area i.e., $E_{nose} > E_{mouth}$. However, this is not necessarily the case during mouth breathing. Because if the subject's nose is clear, there might still exist some involuntary air exchange through the nose even when the subjects try to breathe through the mouth. Since nose breathing and mouth breathing are not mutually exclusive, it might be better to give a ratio indicating how much mouth breathing contributes to the whole air exchange instead of having a binary classifier.

$$p = \frac{E_{\text{mouth}}}{E_{\text{nose}} + E_{\text{mouth}}} \tag{7}$$

3. Experiments

This section describes the experimental setup and protocol for evaluating our proposed classification method. Moreover, the results of our method are presented and discussed.

3.1. Experimental setup and protocol

Eight volunteers (7 males and 1 female aged between 20 and 60) participated in our experiments. Thermography videos were recorded using a FLIR E50 camera[1]. It is a Long Wave Infrared (LWIR) camera featured with thermal sensitivity of better than 0.05K and a spatial resolution of 240×180 pixels. The videos were recorded at 30 frames per second (fps). Furthermore, this camera does calibration/non-uniformity compensation (NUC) every three to four minutes by default in real-time to keep good image quality. The calibration/NUC normally takes around 0.5s, after which the whole image is adjusted by an offset.

All videos were recorded with the subjects lying in the bed in order to simulate a sleeping condition, faces facing right towards the camera. The camera was placed at around 50 cm from the subject's face to cover the view of the whole pillow area such that the face will be in the camera view, even with slight body motions. Each video is four-minute-long. Subjects were asked to breathe with their nose in the first minute and the third minute, mouth in the second and fourth minute while keep stationary and they strictly followed the time protocol. This protocol was used to generate the reference for the benchmark.

During the experiments, we found that some subjects had involuntary air exchange through the nose during mouth breathing which is neither purely nose breathing nor mouth breathing. Therefore, we define a third class of breathing called joint breathing in addition to nose breathing and mouth breathing in which both nose and mouth are used for air exchange. Temporal temperature variations of the nose and mouth area of the three breathing classes are shown in Fig 3. Theoretically, p (as defined in Equation 7) should be 0 during nose breathing and 1 during mouth breathing but considering the sensor noises within the respiratory band, it is impractical to expect p to be exactly 0 or 1. Therefore, we specify a value range for different classes: (1) $p < 0.2$ (nose breathing); (2) $p > 0.8$ (mouth breathing); (3) $0.2 \leq p \leq 0.8$ (joint breathing).

We also manually annotate the windows with three breathing classes based on the temporal variation of the nose area and mouth area in the thermography sequence. And the major difference between the specified protocol and the annotated labels is that some subjects were actually breathing through both nose and mouth when asked to breathe through their mouth.

The accuracy of our method is measured by:

$$acc = \frac{n_{\text{success}}}{n_{\text{success}} + n_{\text{failure}}} \tag{8}$$

where n_{success} and n_{failure} refer to the number of successful classifications and wrong classifications.

3.2. Results and discussion

This section describes the experimental results of eight test subjects. It also discusses the performance difference between mean traces and std traces. The evaluation was implemented and performed using MATLAB[2] (MATLAB R2019b, The MathWorks Inc., Natick, MA, USA).

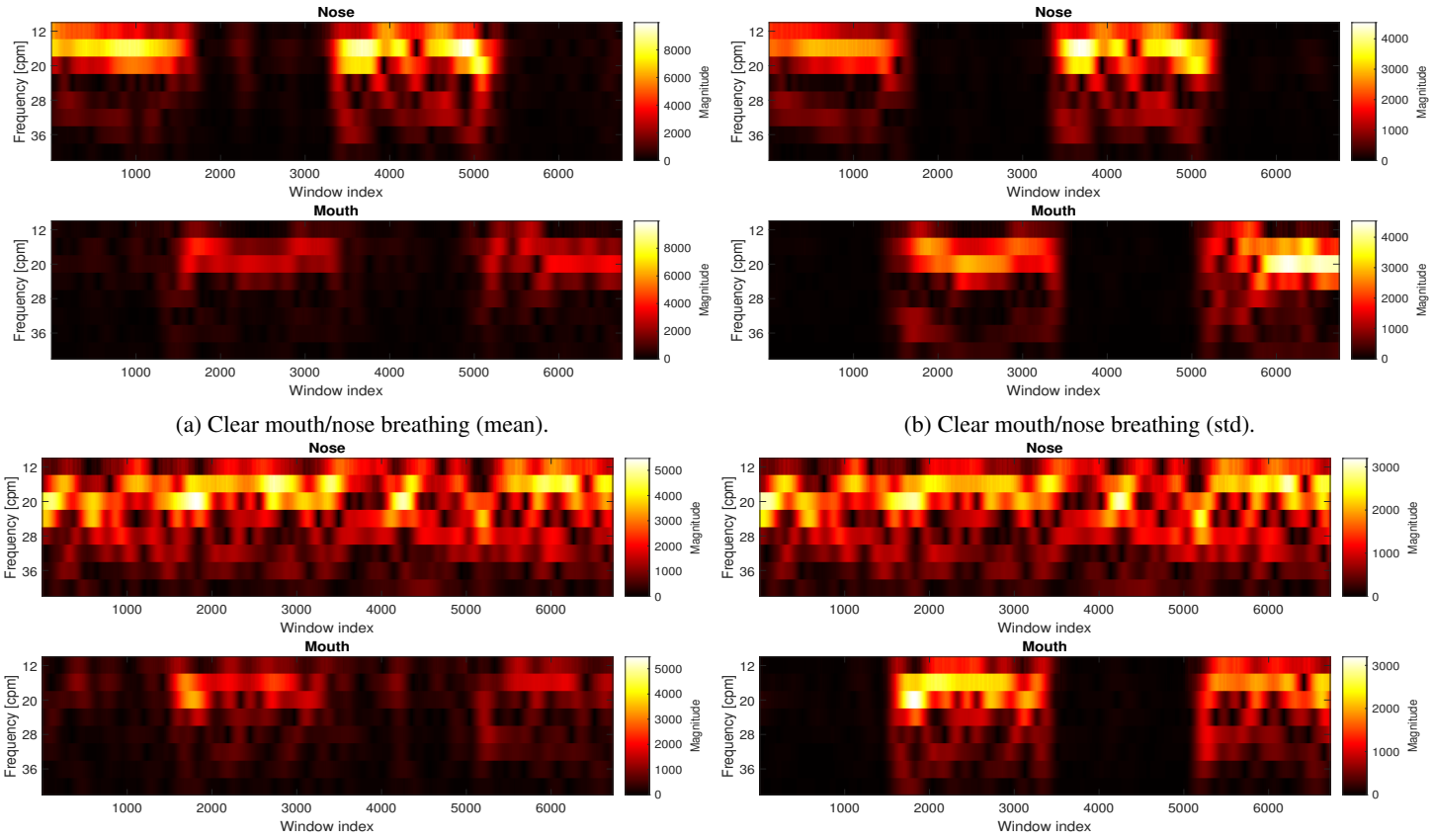
3.2.1 Spectrogram analysis

Fig 2 compares the results obtained by using mean traces and standard deviation traces. Compared to the mean traces of the nose or mouth area which represents the average intensity of the area and captures temporal variations, the standard deviation traces are invariant to the area of ROI due to its characteristics. Moreover, the standard deviation is also immune to the self-calibration of thermal camera which introduces a global shift of the temperature value, as the local standard deviation does not change.

These two experiments also show two representative scenarios. One is clear nose or mouth breathing which is distinguishable as shown in 2a and 2b because the energy level of the nose area is high and the energy level of the mouth area is low during nose breathing (the first and third quarter of the graph) and vice versa. Another one is the joint case where both nose and mouth have airflow during mouth breathing (the second and fourth quarter) as shown in Fig 2c and 2d as we can see that both areas have high energy level.

[1] www.flir.com

[2] www.mathworks.com



(a) Clear mouth/nose breathing (mean).

(b) Clear mouth/nose breathing (std).

(c) Joint mouth-nose breathing (mean).

(d) Joint mouth-nose breathing (std).

Figure 2: Spectrograms of (a) and (c) are obtained from the time windows consisting of average intensity (mean) of nose and mouth areas; of (b) and (d) are obtained from standard deviations (std).

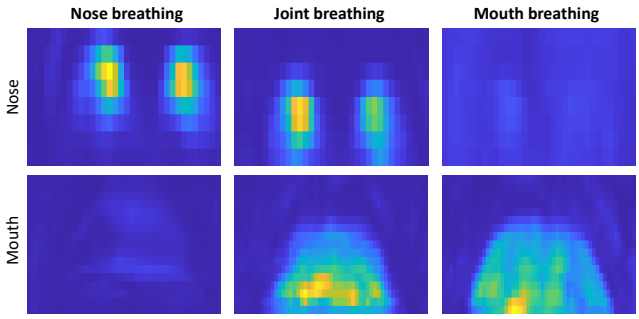


Figure 3: Temporal temperature variations of the nose area and mouth area during nose breathing, joint breathing, and mouth breathing. Only the nose area has large temporal variations (airflow) during nose breathing and only mouth area has large temporal variations during mouth breathing while both areas have large temporal variation in joint breathing.

3.2.2 Classification accuracy

The mouth breathing ratio traces of eight test subjects can be found in Fig 4. As specified by the experimental protocol, the ratio should be low in the first and third minutes and high in the second and fourth minutes. There is also a transition period (in grey shadings) in which windows have both nose breathing and mouth breathing.

	s1	s2	s3	s4	s5	s6	s7	s8	avg
nose	1.00	0.73	1.00	0.84	0.81	1.00	0.74	0.71	0.85
joint	N/A	N/A	N/A	0.50	0.40	0.12	1.00	1.00	0.56
mouth	0.96	1.00	1.00	N/A	N/A	N/A	0.19	N/A	0.87
avg	0.98	0.86	1.00	0.67	0.61	0.56	0.67	0.86	0.78

Table 1: Classification accuracy for eight test subjects (mean).

All eight video recordings are four minutes long, and consist of around 7200 frames, some of which are not available for classification because they contain both nose breathing and mouth breathing frames. Therefore, we have

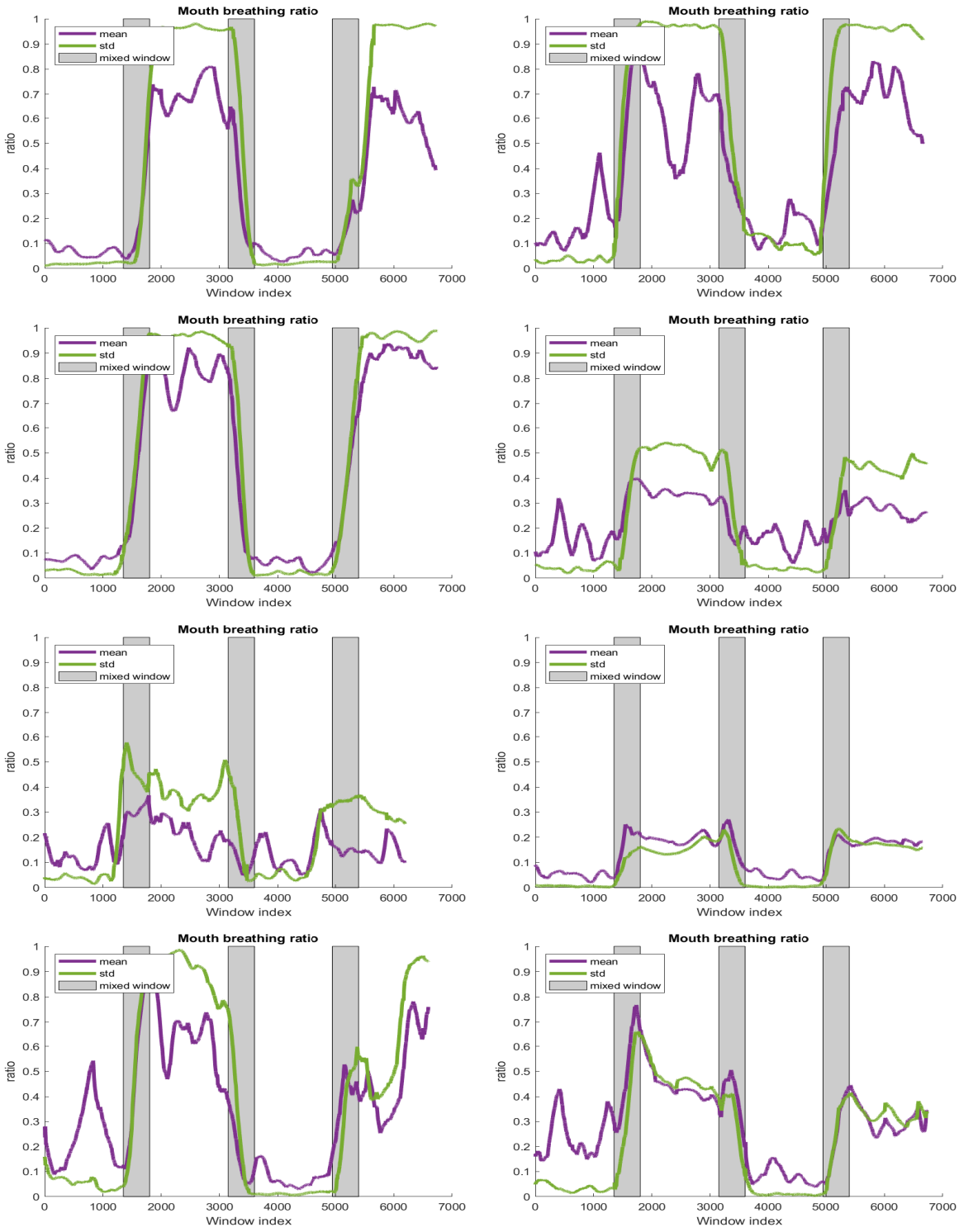


Figure 4: Mouth breathing ratio traces of all eight subjects. It is pure nose breathing frames for windows whose window index is less than 1350, between 3600 and 4950 (the first and third quarter), pure mouth breathing frames for windows whose window index is between 1800 and 3150, between 5400 and 6850 (the second and fourth quarter). Others (shaded area) are mixed windows which consist of both nose breathing frames and mouth breathing frames.

	s1	s2	s3	s4	s5	s6	s7	s8	avg
nose	1.00	1.00	1.00	1.00	0.87	1.00	1.00	1.00	0.98
joint	N/A	N/A	N/A	1.00	1.00	0.02	0.55	1.00	0.73
mouth	0.96	1.00	1.00	N/A	N/A	N/A	0.97	N/A	0.98
avg	0.98	1.00	1.00	1.00	0.93	0.51	0.88	1.00	0.91

Table 2: Classification accuracy for eight test subjects (std).

classification results from around 5400 windows.

The classification accuracy is shown in Table 1 and 2. Overall, our proposed method worked decently on most video recordings and obtained an average accuracy of 91% with std and 78% with mean. The accuracy for joint breathing is a lot lower than the other two with the reason being that subject 6's energy of the mouth area during mouth breathing is too weak to outrace our specified threshold. Our hypothesis is that the subject is not used to breathing through mouth and the ratio of air exchange through the mouth is lower than the threshold.

4. Conclusion

In this paper, a conceptually new measurement, nose-or-mouth breathing classification, has been proposed using thermography, which has clinical/well-being relevance, and can be used as a new feature for sleep monitoring (e.g. early sign for sleep disorder). We also demonstrated this new measurement with an end-to-end image/signal processing flowchart. The results showed that our proposed method achieved a classification accuracy of 91% in ideal lab conditions.

In order to increase relevance in realistic sleeping conditions and improve accuracy, efforts will be necessary to improve nose/mouth localization in non-frontal sleeping poses.

References

[1] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[2] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[3] Sílvio Filipe and Luís A Alexandre. Thermal infrared face segmentation: A new pose invariant method. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 632–639. Springer, 2013.

[4] Healthline. Mouth breathing: Symptoms, complications, and treatments. https://www.healthline.com/health/mouth-breathing. Accessed: 2020-01-13.

[5] Suemy Cioffi Izu, Caroline Harumi Itamoto, Márcia Pradella-Hallinan, Gilberto Ulson Pizarro, Sérgio Tufik, Shirley Pignatari, and Reginaldo Raimundo Fujita. Obstructive sleep apnea syndrome (osas) in mouth breathing children. *Brazilian journal of otorhinolaryngology*, 76(5):552–556, 2010.

[6] Yosh Jefferson. Mouth breathing: adverse effects on facial growth, health, academics, and behavior. *Gen Dent*, 58(1):18–25, 2010.

[7] Marcin Kopaczka, Raphael Kolk, Justus Schock, Felix Burkhard, and Dorit Merhof. A thermal infrared face database with facial landmarks and emotion labels. *IEEE Transactions on Instrumentation and Measurement*, 68(5):1389–1401, 2018.

[8] Marcin Kopaczka, Jan Nestler, and Dorit Merhof. Face detection in thermal infrared images: A comparison of algorithm-and machine-learning-based approaches. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 518–529. Springer, 2017.

[9] Marek Kowalski, Jacek Naruniec, and Tomasz Trzcinski. Deep alignment network: A convolutional neural network for robust face alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 88–97, 2017.

[10] Mariusz Marzec, Robert Koprowski, and Zygmunt Wróbel. Detection of selected face areas on thermograms with elimination of typical problems. *Journal of medical informatics & technologies*, 16:151–160, 2010.

[11] Hawley E Montgomery-Downs and David Gozal. Sleep habits and risk factors for sleep-disordered breathing in infants and young toddlers in louisville, kentucky. *Sleep medicine*, 7(3):211–219, 2006.

[12] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[13] Carina Barbosa Pereira, Xinchi Yu, Michael Czaplik, Rolf Rossaint, Vladimir Blazek, and Steffen Leonhardt. Remote monitoring of breathing dynamics using infrared thermography. *Biomedical optics express*, 6(11):4378–4394, 2015.

[14] Domenick Poster, Shuowen Hu, Nasser Nasrabadi, and Benjamin Riggan. An examination of deep-learning based landmark detection methods on thermal face imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[15] Donald C Rizzo. *Fundamentals of anatomy and physiology*. Cengage Learning, 2015.

[16] Dênis Clay Lopes dos Santos et al. Study of the prevalence of predominantly oral breathing and possible implications for breastfeeding in schoolchildren from são caetano do sul - sp - brazil (original article is in portuguese). *master's dissertation*, 2004.

[17] Wenjin Wang, Albertus C den Brinker, and Gerard De Haan. Full video pulse extraction. *Biomedical Optics Express*, 9(8):3898–3914, 2018.

[18] Yue Wu and Qiang Ji. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 127(2):115–142, 2019.