Low Level Texture Features for Snore Sound Discrimination

Fatih Demir[1], Abdulkadir Sengur[1], Nicholas Cummins[2], Shahin Amiriparian[2,3], Björn Schuller[2,4]

[1]Firat University, Technology Faculty, Electrical and Electronics Engineering Dept., Elazig, Turkey

[2]Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany

[3] Machine Intelligence & Signal Processing Group, Technische Universität München, Germany

[4] GLAM – Group on Language, Audio, and Music, Imperial College London, United Kingdom

***Abstract*— Snoring is often associated with serious health risks such as obstructive sleep apnea and heart disease and may require targeted surgical interventions. In this regard, research into automatically and unobtrusively analysing the site of blockages that cause snore sounds is growing in popularity. Herein, we investigate the use of low level image texture features in classification of four specific types of snore sounds. Specifically, we explore histogram of local binary patterns (LBP) in dense grid of rectangular regions and histogram of oriented gradients (HOG) extracted from colour spectrograms for snore sound characterisation. Support vector machines with homogeneous mapping are used in the classification stage of the proposed method. Various experimental works are carried out with both LBP and HOG descriptors on the INTERSPEECH ComParE 2017 snoring sub-challenge dataset. Results presented indicate that LBP descriptors are better than the HOG descriptors in snore type detection and fusion of the LBP and HOG descriptors produces stronger results than either individual descriptor. Further, when compared to the challenge baseline and state-of-the-art deep spectrum features, our approach achieved relative percentage increases in unweighted average recall of 23.1% and 8.3% respectively.**

Key-words: Snore sound classification, audio spectrograms, low level texture features, local binary patterns, and histogram of oriented gradients.

I. INTRODUCTION

Snore sounds generally occur when the soft palate and uvula structures strike each other and vibrate during breathing. Snoring is considered one of the main symptoms of Obstructive Sleep Apnea (OSA) [1]. OSA is a well-known sleep disorder, which affects approximately 3–7 % of men and 2–5 % of women [1], and can lead to an increased risk of cardiovascular and cerebrovascular diseases [2]. Due in part to such health related effects, the automatic detection of the snore sounds is gaining considerable attention in research and industry. An exemplary application could be an intelligent bed that can recognize the snore-type and adjust the mattress position for termination of the snoring [3].

Snore sound classification has attracted much attention in recent years [4-8]. Spectral and energy features in particular have continually been shown to be highly adept for this task. Cavusoglu et al. proposed an efficient method for discrimination of sleep sounds into snore and non-snore classes based on sub-band energy distribution [4]. Yadollahi et al. proposed an efficient method for discrimination of breath and snore sound segments [5]. The authors extracted various features, namely, energy, zero crossing rate, and formants from the recorded sound signals and employed a Bayesian error based thresholding mechanism for classification. Schmitt et al. used wavelet features, formants, and Mel Frequency Cepstral Coefficients (MFCC) and, in an unsupervised manner, quantized the extracted feature vectors using bag-of-audio-words for snore sound classification [6].

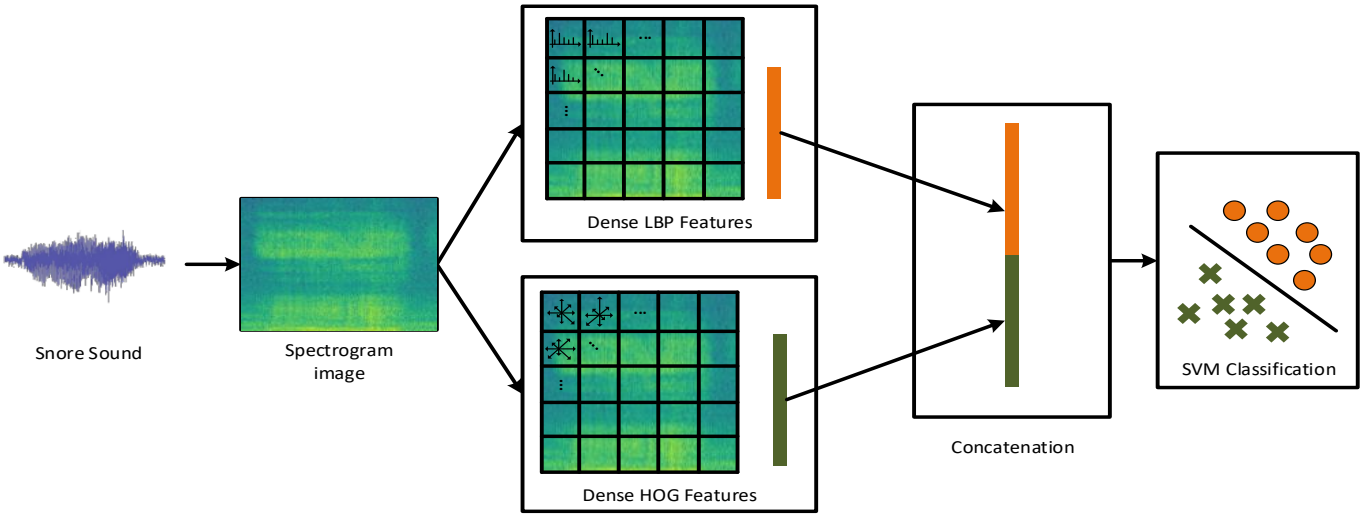
Recently, Amiriparian et al. proposed an efficient approach for snore sound classification [7]. The authors combined deep spectrum features, extracted by passing spectrograms through a convolutional neural network (CNN) pre-trained on ImageNet, and a support vector machine (SVM) classifier in their work. Freitag et al. extend this paradigm and proposed a feature selection for deep spectrum features for improving the accuracy of the snore sound classification [8]. To this end, the authors employed the particle swarm optimization for feature selection.

From the reviewed literature, it can be seen that various approaches have been proposed for efficient classification of the snore sounds. The proposed methods have generally employed the time domain features, while a few notable works have concentrated on time-frequency (t-f) representation (spectrogram images) based features. Spectrograms contain rich t-f information; this property makes them convenient for various applications such as electroencephalogram (EEG) and Electromyography (EMG) classifications [9, 10]. In addition, features derived from spectrograms have been found to be effective for snore sound analysis [8].

In this regard, this work proposes a methodology for the efficient classification of the upper airway obstructions from the recorded snore sounds. The proposed approach initially converts the recorded snore sounds to the colour images based on the spectrogram. Local binary pattern (LBP) and histogram of oriented gradients (HOG) approaches are then used to extract distinctive features from the spectrogram images. The extracted LBP and HOG features are then concatenated and classified with a SVM classifier. To the best of the authors' knowledge, this is the first time that low level image features have been used for audio based snore sound classification.

Our experiments are conducted on the Munich-Passau Snore Sound Corpus (MPSSC) from the INTERSPEECH ComParE 2017 snoring challenge [11]. We compare the obtained results with the challenge baseline and various state-of-the-art-methods. These indicate that our approach clearly improves the classification accuracy when both the baseline and state-of-the-art-methods results are considered [8].

Figure 1. The illustration of the proposed method. Spectrogram images are generated from the input snore sounds and saved as colour images with Matlab. The obtained colour images are then used as input to LBP and HOG feature extractors. Dense LBP and HOG features are extracted and then concatenated. Finally, support vector machine (SVM) classifier is used for the classification.



II. PROPOSED APPROACH

Our system is composed of three main components (cf. Fig. 1): (i) The construction of the spectrogram images from the recorded snore samples, (ii) LBP and HOG based feature extraction and, (iii) a SVM classifier.

A. Spectrogram Images

A spectrogram is a t-f imaging technique that is represented by the squared magnitude of short–time-Fourier-transform (STFT). The STFT divides an input signal into short pieces and employs the same observation time-resolution in all frequency-bandwidths; i.e., its time-resolution and the frequency-resolution is uniform. Matlab was used to construct the spectrogram images. We use Hamming windows of width 32 ms, and overlap 8 ms, and compute the power spectral density on the dB power scale. The spectrogram images are saved with a viridis colour map, which is a perceptually uniform sequential colour map varying from blue to green to yellow. The constructed spectrogram images have an intermediate size of 875×656 pixels.

B. Local Binary Pattern (LBP) Features

The LBP process converts an image into an array of integer labels [12]. In other words, it encodes the local structure around each pixel which constructs a small scale appearance of the considered image. In the LBP procedure, the centre pixel in a 3×3 neighbourhood is compared with its eight neighbours by subtracting values. The obtained negative values are labelled with 0 and the rest with 1. An eight bits binary number is obtained by concatenating all these binary codes in a clockwise direction starting from the top-left one and its corresponding decimal value is used for labelling.

C. Histogram of Oriented Gradients (HOG) Extraction

HOG is an efficient texture descriptor; it starts its process by dividing an input image into a dense grid of rectangular regions [13]. For each region of interest, the orientations of the gradients are computed. A histogram is constructed for each region based on the accumulation of the weighted votes of the gradient magnitudes of each pixel. Generally, for local histogram generation, an 8×8 window, nine bins and 0-180 degrees orientation range are considered. Furthermore, the rectangular regions are grouped into blocks and all region histograms are normalized. Because of overlapping, the same region can be differently normalized in several blocks. The descriptor is calculated using all overlapping blocks from the image detection window.

III. EXPERIMENTS

A. Dataset

The Munich-Passau Snore Sound Corpus (MPSSC), made publicly available through the INTERSPEECH 2017 ComParE challenge was used to obtain experimental results [18]. The dataset contains 828 snore audio files in total that are grouped into four categories (Velum, Oropharyngeal lateral walls, Tongue base and Epiglottis). The sampling rate of the snore audio files is 16 kHz and each file has various sample lengths. For the challenge, the snore audio files were partitioned into training (282), development (283), and test (263) partitions.

B. Experimental settings and results

We use Matlab R2014b on a computer having an Intel Core i7-4810 CPU and 32 GB memory. In spectrogram creation of the snore audio files, the number of the FFT is chosen as 512. The parameters for the spectrogram are determined heuristically during initial experimental works (results not given). The saved original spectrogram images have an initial size of 875×656 and are then resized to 227×227 to reduce the time complexity associated with feature extraction, furthermore, this allows a more even comparison with results presented in [7], who also resized their spectrogram to this dimensionality.

We opt to extract dense LBP features where the input image is divided into a grid of rectangular cells. For each cell,

Table 1. Results for LBP features on the snore sub-challenge using a linear SVM on different colour channels and their combinations. Unweighted Average Recall (UAR %) is used as measure and *c* is optimized on the development partition.

Feature type	Feature dimension	Colour information	*C* parameter of SVM	Development (%)	Test (%)
LBP features	2891	Grey	0.1	33.0	-
	2891	R channel	10	34.2	-
	2891	G channel	0.01	34.5	-
	2891	B channel	1	34.2	-
	5782	R, G channels	0.1	33.1	-
	5782	R, B channels	1	35.4	62.7
	5782	G, B channel	10	33.9	-
	8673	R, G and B channels	0.01	34.6	64.4

Table 2. Results for HOG features on the snore sub-challenge using a linear SVM on different colour channels and their combinations. Unweighted Average Recall (UAR %) is used as measure and *c* is optimized on the development partition.

Feature type	Feature dimension	Colour information	*C* parameter of SVM	Development (%)	Test (%)
HOG features	1296	Grey	1	37.7	60.2
	1296	R channel	0.1	36.1	-
	1296	G channel	0.01	37.1	-
	1296	B channel	0.1	34.4	-
	2592	R, G channels	0.1	37.3	-
	2592	R, B channels	100	35.5	-
	2592	G, B channel	0.01	35.0	-
	3888	R, G and B channels	0.1	35.0	59.7

a 59 dimensional histogram is computed, with all histograms vector (2891 dimensional). For the generation of the LBP histogram, the following parameters are selected heuristically; the number of neighbours is selected as 8, the radius of circular pattern to select neighbours is assigned to 1, and a 32×32 range is chosen as the cell size. In addition, for HOG features extraction, the related parameters are set as follows; the cell size is set to 32×32, the block size is set to 2×2, and the number of bins is chosen as 9. Thus, a 1296 dimensional HOG feature is computed for each input image. Note that, the parameters for the LBP and HOG features are also determined heuristically during initial experimental works (results not given).

The SVM classifier with homogenous mapping and the LIBLINEAR library with the L2-regularised L2-loss dual solver is considered because of its robustness to smaller amounts of training data [14]. The SVM parameter *C* is searched in the range of $[10^{-3}, 10^{-2}, \ldots, 10^{3}]$. As per the ComParE challenge, the Unweighted Average Recall (UAR) is used to evaluate the performance of the proposed method. During initial experimental works, we investigated the performance of the grey level and red (R), green (G) and blue (B) channels and the colour channel combinations. The results of our experiments across all colour channels and their combinations are given in Tables 1, 2 and 3.

IV. Results

While Table 1 shows the LBP features classification performance, Table 2 and Table 3 show the HOG descriptors achievements and concatenated LBP+HOG features achievements, respectively. As shown in Table 1, the best result for the LBP features on the development partition are obtained with the R and B channels in combination.

The UAR is 35.4% and the C value is 1; the corresponding test set UAR for this set-up is 62.4%. In addition, a slightly stronger result for the test partition is obtained with the R, G and B channels in combination, where the UAR and C values are 64.4% and 0.01, respectively. We speculate that the stronger test set results are due to the increase in available data to train the system when the training and development partitions are combined. In general, the results indicate that a combination of colour channels produces improved results over single channel colour systems.

The HOG descriptor's achievements for grey scale, single colour channels and combination of the colour channels are given in Table 2. For the development partition, the strongest result is obtained for grey scale where the C parameter is 1; the calculated UAR is 37.7%. In addition, the highest UAR value of 60.2% for the test partition is obtained for grey-scale HOG features. Moreover, the second best UAR value 59.5% for the test partition is obtained for the R, G and B colour channels in combination. Given the trend observed in the LBP features, this result is surprising; we expected the colour channels combination achievements to be higher than the single channels achievements.

However, it is worth mentioning that for the development partition, the G channel, and R and G channels combination produce UAR values quite close to the highest UAR value i.e., 37.1% and 37.3%, respectively. The obtained results computed with the concatenated LPB and HOG descriptors are shown in Table 3. One important output from Table 3 is that while the fusion of the features does not lead to large gains in the development partition, concatenation considerably improves the classification results for the test partition. The highest UAR (37.8%) for the development partition is obtained with the G and B channels in combination, where the C parameter is 0.1.

Table 3. Results for concatenated LBP+HOG features on the snore sub-challenge using a linear SVM on different colour channels and their combinations. UNWEIGHTED AVERAGE RECALL (UAR %) is used as measure and *c* is optimized on the development partition.

Feature type	Feature dimension	Colour information	*C* parameter of SVM	Development (%)	Test (%)
LBP + HOG features	4187	Gray	0.1	36.4	-
	4187	R channels	10	35.0	-
	4187	G channels	0.1	36.1	-
	4187	B channels	0.01	36.7	-
	8374	R, G channels	0.01	36.6	-
	8374	R, B channels	0.1	36.8	69.0
	8374	G, B channel	0.1	37.8	72.6
	12561	R, G and B channels	0.01	35.6	-

Table 4. Comparison of our proposed method with baseline results and other state-of-the art methods

Method	Development %	Test %
Baseline CNN&LSTM [11]	40.3	40.3
Baseline functionals [11]	40.6	58.5
Deep Spectrum [7]	44.8	67.0
Deep Spectrum & CSO [8]	57.6	66.5
Snore Challenge Winners [15]	NA	64.2
Proposed method	37.0	72.6

In addition, the highest UAR value (72.6%) is obtained with the G and B channels in combination for the test partition, where C is 0.1. This result is higher than the 64.2% UAR achieved by the winners of the ComParE snore sound challenge [15] and the best known (to date) test set UAR, 67.0%, achieved with deep spectrum features (cf. Table 4).

In general, these results indicate that the LBP descriptors produce stronger results than the HOG descriptors, and the concatenated LBP and HOG features further improve the classification performance. This result provides further evidence in support of our earlier speculation that the SVM classifier requires a greater number of LBP and HOG descriptors training sample to adequately model the related input-output relation.

V. CONCLUSIONS

In this paper, we investigated the applicability of low-level texture image descriptors for audio-based snore sound classification. To this end, the popular LBP and HOG descriptors were considered. The snore audio files were initially transformed to the colour spectrogram images. Then, grey scale images, each colour channels and colour channels in combinations were considered in our experiments.

The obtained results revealed that the LBP descriptors were stronger than the HOG descriptors. In addition, combination of the colour channels descriptors also yielded better results than the single channel descriptors. Moreover, concatenation of the LBP and HOG descriptors achieved the highest score. To the best of the authors' knowledge, the obtained results are the highest reported to date on the INTERSPEECH 2017 ComParE Snoring dataset.

Based on our observation that a greater number of LBP and HOG features are required to train a more effective classifier, in the future work, we are planning to explore the advantages of augmenting the training and development partitions with artificially generated data.

VI. REFERENCES

[1] M. S. Aldrich, Sleep medicine. Oxford University Press, 1999.

[2] O. Parra, A. Arboix, J. Montserrat, L. Quinto, S. Bechich, and L. Garcia-Eroles, "Sleep-related breathing disorders: impact on mortality of cerebrovascular disease," *European Respiratory Journal*, vol. 24, no. 2, pp. 267–272, 2004.

[3] F. Crivelli, E. Wilhelm, R. van Sluijs, R. Riener, "Actuated bed for a closed loop anti-snoring therapy," *In Rehabilitation Robotics (ICORR), 2017 International Conf. on*, London, UK: IEEE, 2017, pp. 823-827.

[4] M. Cavusoglu, M. Kamasak, O. Erogul., T. Ciloglu, Y. Serinagaoglu., T. Akcam, "An efficient method for snore/nonsnore classification of sleep sounds,"*Physiological Measurement*, vol. 28, no. 8, pp. 841-853, 2007.

[5] A. Yadollahi, Z. Moussavi., "Automatic breath and snore sounds classification from tracheal and ambient sounds recordings," *Medical Engineering & Physics*, vol. 32, no. 9, pp. 985-990, 2010.

[6] M. Schmitt, C. Janott, V. Pandit, K. Qian, W. Hemmert, C. Heiser, B. Schuller. "A Bag-of-Audio-Words Approach for Snore Sounds' Excitation Localisation." *Speech Communication; 12. ITG Symposium; Proceedings of*. Paderborn, Germany VDE, 2016, pp 1-5.

[7] S. Amiriparian, M. Gerczuk, S. Ottl, N. Cummins, M. Freitag, S. Pugachevskiy, A. Baird, B. Schuller, "Snore sound classification using image-based deep spectrum features." *in Proceedings INTERSPEECH 2017*, Stockholm, Sweden: ISCA, August 2017, pp. 3512-3516

[8] M. Freitag, S. Amiriparian, N. Cummins, M. Gerczuk, B. Schuller, "An 'end-to-evolution'hybrid approach for snore sound classification." *in Proceedings INTERSPEECH 2017*, Stockholm, Sweden: ISCA, August 2017, pp. 3507-3511

[9] A. Şengür, Y. Guo, Y. Akbulut, "Time–frequency texture descriptors of EEG signals for efficient detection of epileptic seizure," *Brain Informatics*, vol. 3, no. 2, pp. 101-108, June 2016.

[10] A. Sengur, M. Gedikpinar, Y. Akbulut, E. Deniz, V. Bajaj, Y. Guo, "DeepEMGNet: an application for efficient discrimination of ALS and normal EMG signals," *Mechatronics 2017*. Springer, pp 619-625, 2017.

[11] B. Schuller, S. Steidl, A. Batliner, E. Bergelson, J. Krajewski, C. Janott, A. Amatuni, M. Casillas, A. Seidl, M. Soderstrom, A. Warlaumont, G. Hidalgo, S. Schnieder, C. Heiser, W. Hohenhorst, M. Herzog, M. Schmitt, K. Qian, Y. Zhang, G. Trigeorgis, P. Tzirakis, and S. Zafeiriou, "The INTERSPEECH 2017 Computational Paralinguistics Challenge: Addressee, Cold & Snoring," *in Proceedings INTERSPEECH 2017*, Stockholm, Sweden: ISCA, August 2017, pp. 3442-3446

[12] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002.

[13] N. Dalal, B. Triggs., "Histograms of oriented gradients for human detection," *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognition*, San Diego, CA, USA: IEEE, June 2005, pp. 886–893.

[14] R.-E. Fan, K.-W., Chang, C.-J., Hsieh, X.-R. Wang, C.-J. Lin., "LIBLINEAR: A library for large linear classification," Journal of Machine Learning Research, vol. 9, pp. 1871-1874, Aug. 2008.

[15] H. Kaya, A. A. Karpov, "Introducing Weighted Kernel Classifiers for Handling Imbalanced Paralinguistic Corpora: Snoring, Addressee and Cold. Proc.," *in Proceedings INTERSPEECH 2017*, Stockholm, Sweden: ISCA, August 2017, pp. 3527-3531