Using spectro-temporal features for Environmental Sounds

Souli sameh #1, Zied lachiri #*2
#Signal, Image and pattern recognition research unit
Dept. of Genie Electrique, ENIT
BP 37, 1002, Le Belvédère, Tunisia
1 soulisameh@yahoo.fr
**Dept. of Physique and Instrumentation, INSAT*
BP 676, 1080, Centre Urbain, Tunisia
2 ziedlachiri@enit.rnu.tn

Abstract

The paper presents the task of recognizing environmental sounds for audio surveillance and security applications.

A various characteristics have been proposed for audio classification, including the popular Mel-frequency cepstral coefficients (MFCCs) which give a description of the audio spectral shape. However, it exist some temporal-domain features. These last have been developed to characterize the audio signals. Here, we make an empirical feature analysis for environmental sounds classification and propose to use the log-Gabor-filters algorithm to obtain effective time-frequency characteristics.

The Log-Gabor filters-based method utilizes time-frequency decomposition for feature extraction, resulting in a flexible and physically interpretable set of features.

The Log-Gabor filters-based feature is adopted to supplement the MFCC features to yield higher classification accuracy for environmental sounds.

Extensive experiments are performed to prove the effectiveness of these joint features for environmental sound recognition. Besides, we provide empirical results showing that our method is robust for audio surveillance Applications.

Keywords: Environmental sounds, MFCC, Log-Gabor filters, Spectrogram, SVM Multiclass.

1 Introduction

Many previous works have focused on the recognition of speech and music while research on environmental sounds recognition has received little attention. Some efforts have been emerged toward systems which investigate environmental classification [1-2].

Besides, the courant life sounds are very versatile, that composed of sounds generated in domestic, business, and outdoor environments.

The high variability of sounds makes such model difficult to manipulate, the majority of works concentrate on specific classes of sounds.

There is system that is able to classify environmental sounds. This system possesses a great importance for surveillance and security applications [2]. The aim is the identification of some current life sounds class. Among the eventual applications [3-4-5-6], we quote: the cars classification according to their noise, the fire arms sounds identification to warn the police, the distress sounds identification for the remote monitoring systems and medical security [7].

In this paper, the system elaborated is adapted for the classification of a few number of environmental sounds classes and is interested by a sound-based surveillance application.

In standard sound classification methods [8-9], the classification of a sound is usually composed of two phases.

First, a set of features is generated using various techniques to characterizing the signal to be classified.

Then, for these feature vectors, a classifier is used to assign a pattern to a class. The select of proper features is necessary to obtain an effective system performance.

In this work, our focus is to characterize the environmental sounds types. Generally, audio signals have been characterized by the popular Mel-frequency cepstral coefficients (MFCCs) or time-frequency representations like the wavelet transform.

In the literature, the filter bank used for MFCC computation possesses some significant properties of the human auditory system. The use of MFCCs for structured sounds in particular speech and music have been obtained a good performance to characterizing signal, but their performance degrades in the presence of noise.

We can conclude also that MFCCs are not capable to analyzing signals that possessed a flat spectrum [21].

Most of environmental sounds have a broad flat spectrum that may not be effectively modeled by MFCCs.

Courant life sounds form a large and diverse variety of sounds, like explosions and gunshots which have a strong temporal domain signatures, these sounds have a broad flat spectrum which are sometimes not effective to model by MFCCs.

In this work, we propose to use the Log-Gabor filters (LGF) in addition to MFCCs coefficients to analyze environmental sounds. Log-Gabor filters (LGF) offer a way to extract time-frequency domain features that can classify sounds. They provides an excellent simultaneous localization of spatial and frequency information [10]. The process contains finding the decomposition of a signal from spectro-temporal components, which would yield the best set of functions to obtain an approximate representation. The log-Gabor filters coefficients contain relevant and effective information. They consist in signal decomposition into spectro- temporal atoms, which are efficient to form an approximate representation.

The log-Gabor filter has been used in a variety of applications, such as speech detection [11] and Stress emotion classification [10]. Log-Gabor filter has also been used in image genre classification [12].

In [13] Gabor filters have been proposed, as the face identification techniques. Other works have used Gabor wavelets in the elastic comparison graphs [14] and in the correlation of Gabor filter representations [15].

Other studies have used Gabor filters for the fingerprint identification [16], for the segmentation of the texture [17], for identification of the iris [18] and identification of face [19].

In our proposed approach, the log-Gabor filter is used for feature extraction in the context of environmental sound [20]. We investigate a combination of features and ensure an empirical evaluation on ten environment classes.

It is demonstrated that the most frequently-used features do not always efficient with environmental sounds while the Log-Gabor filters-based features can be added to frequency domain features (MFCC) to produce higher classification accuracy for courant life sounds.

This paper is organized as follows. Some interesting previous work is discussed in Section 2. Section 3 presents a review of different audio feature extraction methods. The log-Gabor filter algorithm is described and the combination of the log-Gabor filters based features and MFCCs is presented in Section 4. Section 5 describes experimental evaluation of selected features. Finally conclusions and perspectives are presented in Section 5.

2 Background Review

A major problem in construction of an automatic audio classification system is the choice of signal characteristics which may lead an effective discrimination between various environmental sounds.

Unlike music or speech, generally environmental sounds possess unstructured data including of contributions from a variety of sources. In this case, it is difficult to constitute a generalization to quantify unstructured data.

Because of the variety and diversity sound, it exist many features that can be used, to describe environmental sound.

Generally, acoustic characteristics can be divided into two domains: time-domain and frequency-domain.

In order to construct a robust classification system, the suitable choice of these features is essential.

For each type of environmental sound, it exist some underlying structures, so we used log-Gabor filters to discover them [20].

Various types of courant life sounds possess their own unique characteristics, which enables to notice that the decomposition into sets of basis vectors to be noticeably different from one another.

We have demonstrated in [20] that the log-Gabor filters constitute an efficient way of selecting a small group of basis vectors that promotes the production of meaningful features in order to characterize an environmental sound [21].

The log-Gabor filters algorithm was originally applied to reassigned spectrogram of environmental sounds [22].We are used time-frequency representations in particular sound spectrogram, which offers new opportunities for promising parameterization [23].

The advantage of the time-frequency representation is the ability to bring out the useful structure of each type of sound [10].

In order to improve the readability and eliminate interference of spectrogram we proposed to apply the reassignment method. This method relies on the intervention of an adequate field of vectors which moves the values of the time-frequency distribution so that at the end, reading becomes simplified [24].

The reassignment approach refocuses spectrogram energy components and corrects the low concentration time-frequency [22].

- Log-Gabor filters:

The log-Gabor filters consist in signal decomposition into spectro- temporal atoms. They have many useful and important properties, in particular the capacity to decompose an image into its underlying dominant spectro-temporal components [25-26]. The log-Gabor function in the frequency domain can be presented by the transfer function $\boldsymbol{G(r,\theta)}$ with polar coordinates [10]:

$$G(r,\theta) = G_{radial}(r).G_{angular}\ (r) \qquad (1)$$

Where $G_{radial}(r) = e^{-\log(r/f_0)^2/2\sigma_r^2}$, is the frequency response of the radial component and $G_{angular}(r) = exp\,(-(\theta/\theta_0)^2/2\sigma_\theta^2)$, represents the frequency response of the angular filter component.

We note that (r,θ) are the polar coordinates, f_0 represents the central filter frequency, θ_0 is the orientation angle,

σ_r and σ_θ represent the scale bandwidth and angular bandwidth respectively.

The log-Gabor feature representation $|S(x,y)|_{m,n}$ of a magnitude spectrogram $s(x,y)$ was calculated as a convolution operation performed separately for the real and imaginary part of the log-Gabor filters:

$$Re(S(x,y))_{m,n} = s(x,y) * Re\big(G(r_m,\theta_n)\big) \qquad (2)$$

$$Im(S(x,y))_{m,n} = s(x,y) * Im\big(G(r_m,\theta_n)\big) \qquad (3)$$

(x,y) represent the time and frequency coordinates of a spectrogram, and $m = 1,\dots,N_r = 2$ and $n = 1,\dots,N_\theta = 6$ where N_r devotes the scale number and N_θ the orientation number. This was followed by the magnitude calculation for the filter bank outputs:

$$|S(x,y)| = \sqrt{\left(Re\big(S(x,y)\big)_{m,n}\right)^2 + Im(S(x,y))_{m,n}} \qquad (4)$$

The feature vectors are calculated by an averaged operation for each 12 log-Gabor filter appropriate. The purpose being to obtain a single output array [10]:

$$\left|\hat{S}(x,y)\right| = \frac{1}{N_r\,N_\theta}\sum_{\substack{m=1\\n=1}}^{N_r,N_\theta} |S(x,y)|_{m,n} \qquad (5)$$

We processed three approaches. In the first approach, a reassigned spectrogram is generated from sound. Next, it goes through single log-Gabor filter extraction with 2 scales $(1,2)$ and 6 orientations $(1,2,3,4,5,6)$. Then, we apply mutual information in order to get an optimal feature. This feature is finally used in the classification (figure 1).

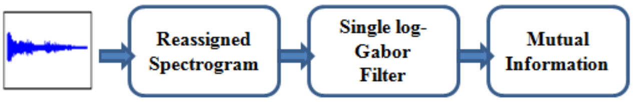

Figure 1 Feature extraction using single log-Gabor filter

The second approach consists of the same steps as first one, but with an averaged 12 log-Gabor filters $\{G_{11}, G_{12}, \dots, G_{16}, G_{21}, \dots, G_{25}, G_{26}\}$, instead of single log-Gabor filter (figure 2).

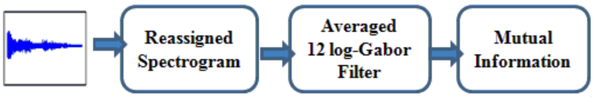

Figure 2 Feature extraction using 12 log-Gabor filters

In the third approach the idea is to segment each spectrogram into 3 patches. Intuitively, for each patch, averaged 12 log-Gabor filters are calculated. After that we apply a mutual information selection to pass then in the classifier. In the classification phase, we use SVM, in One-Against-One configuration with the Gaussian kernel (figure 3). For more information we can see [22].

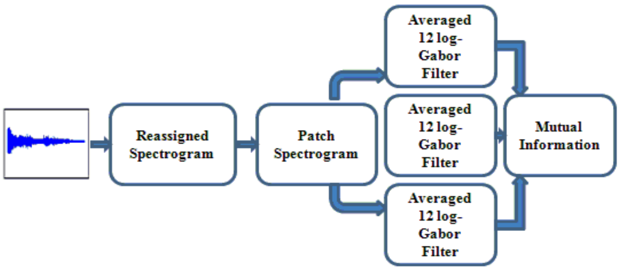
Reassigned Spectrogram
Patch Spectrogram
Averaged 12 log-Gabor Filter
Averaged 12 log-Gabor Filter
Averaged 12 log-Gabor Filter
Mutual Information

Figure 3 Feature extraction using 3 spectrogram patches with 12 log-Gabor filters

3 Features Extraction with Log-Gabor Filter and MFCCs

In order to have a better classification, the good choice of feature is essential. The used features should be robust, stable, and physically interpretable.

In this paper we will show that the use of log-Gabor filters added to MFCCs is very efficient for classification system.

The advantages of the combination of Log-Gabor filters and MFCCs are the ability to capture the inherent structure within each type of environmental sounds.

Our aim is to use log-Gabor filters added to MFCCs as a tool for feature extraction for classification. Nevertheless, the combination of Log-Gabor filters and MFCCs provides an excellent improvement in the recognition results compared to results obtained when using only log-Gabor filters.

We chose to use the concatenation of 12 log-Gabor filters. This choice is justified in our previous work [20], where we have shown that the concatenation of 12 log-Gabor filters is obtained the best classification rate.

In the literature, we remarked that among the suitable audio features for combination are the Mel-frequency cepstral coefficients (MFCC).

In [21], the addition of MFCC to Matching Pursuit achieved the best classification rate compared to other audio features such as the short-time energy, the zero crossing rate and the spectral flux.

We remarked also in [27] that the use of MFCC in addition with temporal and wavelet based features improve the system performance.

Log-Gabor filters are parameterized in frequency and orientation. They have the advantage of extracting localized and oriented frequency information. [28], [29]. They provide an excellent simultaneous spatial and frequency localization of information. They have several important properties, particularly the ability to decompose a spectrogram into its dominant spectral and spatial components [30].

However, we chose the log Gabor filter to extract relevant descriptors for two reasons. First, the log-Gabor functions don't have continuous component, which helps to improve the contrast of edges, and the borders of spectrograms. Second, the transfer function of the log-Gabor function has a long tail on the extremity of high frequency, which allows us to obtain wide spectral information with localized spatial extent and contributes, thus, to preserve the true structures of edges of spectrograms [29].

The important aspect of the function of log-Gabor is that, contrary to the Gabor function, the frequency response of the log Gabor is symmetric on a logarithmic axis.

Log-Gabor filters can be constructed with a given bandwidth. This bandwidth can be optimized to produce a filter with minimal spatial extent.

It was shown that the functions of log-Gabor has extensive queues at high frequency extremities should be able to encode spectrogram more effectively through better representation of high frequency components.

4 Experimental Evaluation

4.1. Experimental Setup

We examined the performance of the features and make an experimental evaluation on ten different types of current life sounds as shown in Table 1.

The corpus sound samples used derived from different sound libraries available [31-32]. Otherwise, using several sound collections is important and very necessary to create a representative, large, and enough diverse databases.

The used database contains impulsive and harmonic sounds for example phone rings (Pr) and children voices (Cv). All signals have a resolution of 16 bits and a sampling frequency of 44100 Hz that is characterized by a good temporal resolution and a wide frequency band, which are both necessary to cover harmonic as well as impulsive sounds.

Table 1 Classes of sounds and number of samples in the database used for performance evaluation

Classes	Train	Test	Total
Door slams (Ds)	208	104	312
Explosions (Ep)	38	18	56
Glass breaking (Gb)	38	18	56
Dog barks (Db)	32	16	48
Phone rings (Pr)	32	16	48
Children voices (Cv)	54	26	80
Gunshots (Gs)	150	74	224
Human screams (Hs)	48	24	72
Machines (Mc)	38	18	56
Cymbals (Cy)	32	16	48
Total	670	330	1000

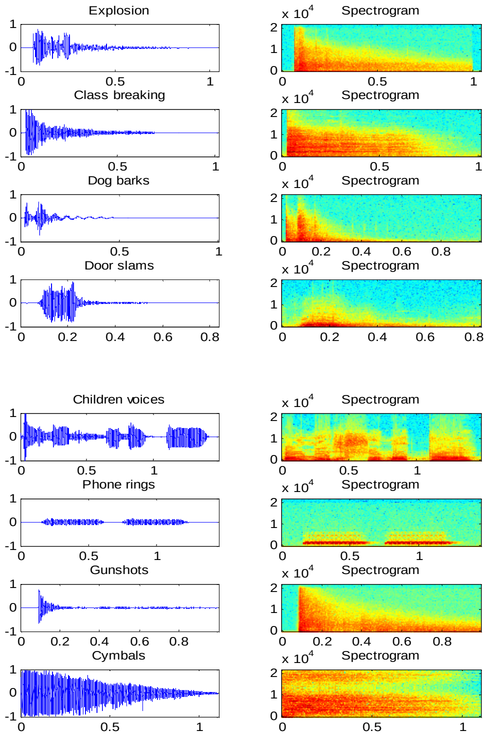

Figure 4 Audio waveform and Spectrograms of 8 classes environmental sound.

The ten environment types considered were: Door slams, Explosions, Glass breaking, Dog barks, Phone rings, Children voices, Gunshots, Human screams, Machines, Cymbals.

In addition, we remark the presence of some classes sound very similar to human listeners such as explosions (Ep) are pretty similar to gunshots (Gs) (figure 4), hence, it is sometimes not obvious to discriminate between them. They are deliberately differentiated to test capacity of the system in separating very similar classes of sounds.

A type of sounds is required by the application, sounds are non-still, mainly of short durations, mainly impulsive audio signals, and presenting a big diversity intra-classes and a lot of similarity inter-classes. Most of the impulsive signals introduced into the base have duration of 1s, but some sounds possess much superior durations which can achieve 6s (for certain samples of explosions and the Human screams).

We examined the performance of 12 log-Gabor filters features, MFCC (12), a concatenation of the log-Gabor filters features and MFCCs.

We adopted the Gaussian Mixture Model (GMM) and the Support Vector Machines (SVM) and in the classification phase.

We begin with GMMs which for each data class was modeled as a mixture of several Gaussian clusters. The conditional probabilities were computed with formula below:

$$p(x \setminus X_k) = \sum_{j=1}^{m_k} p(x \setminus j)P(j)$$

where X_k is the data points for each class, m_k is the number of components, $P(j)$ is the prior probability, and $p(x \setminus j)$ is the mixture component density. Then, the EM algorithm [33] was generated to obtain the maximum likelihood parameters of each class.

We used also a Support Vector Machine, in One-against-One and One-against-All configuration [34].

The idea is to employ a kernel function $K(x_i, x_j)$, where $K(x_i, x_j)$ satisfies the Mercer conditions [35]. We chose a Gaussian RBF kernel:

$$k(x, x') = exp\left[\frac{-\|x-x'\|^2}{2\sigma^2}\right]. \qquad (6)$$

Where $\|.\|$ indicates the Euclidean norm in $\Re^d$.

Ω allows to perform a mapping of a large space in which the linear separation of data is possible [36].

$$\Omega: \Re^d \longrightarrow H$$
$$(x_i, x_j) \longmapsto \Omega(x_i)\Omega(x_j) = k(x_i, x_j) \qquad (7)$$

The H space reproduces kernel Hilbert space (RKHS) of functions. Thus, the dual problem is presented by a Lagrangian formulation as follows:

$$\max W(\alpha) = \sum_{i=0}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j k(x_i, x_j)|_{i=1,\dots,m} \qquad (22)$$

Under the following constraints:

$$\sum_{i=1}^{m} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C. \qquad (8)$$

The α_i are called Lagrange multipliers and c is a regularization parameter which is used to allow classification errors. The decision function will be formulated as follows:

$$f(x) = sgn(\sum_{i=1}^{m} \alpha_i \, y_i k(x, x_i) + b) \qquad (9)$$

We adopted One-against-One and One-against-All approaches [37].

4.2. Experimental Results

In this section we begin by the presentation of classification results which obtained when using only one feature in the feature vector.

The MFCC feature [38] is computed from each frame of the reassigned spectrogram. We used the Hamming analysis window, with length 25 ms and 50% overlap.

Concerning the computation of log-Gabor filters, a concatenation of 12 filters was applied to the reassigned spectrogram. The 12 log Gabor filters are derived from 2 scales and six orientations. In order to improve the time-frequency representation and eliminate interferences reassigned spectrogram is used [22].

Evaluations of the M-SVM-based system using a Gaussian RBF kernel with individual features are compared to the results obtained by the GMM-based classifier.

Table II contains the results. We performed a comparison using GMMs, M-SVM(1-vs-1), and M-SVM(1-vs-all).

According to the results, presented in table 2, the 1-vs-1 classifier performs better than 1-vs-all and GMM classifiers.

We remark also that none of the individual features are able to attain very high performance. In this case, the use of features combination is a solution, as presented in the next subsection.

Table 2 Recognition Rates Using Various Features Applied to GMMs, and M-SVMs- Based Classifiers

Features	*Recognition Rate %*		
	GMM	*M-SVM(1-vs-1)*	*M-SVM(1-vs-all)*
12MFCCs	*81.52*	*83.87*	*81.82*
12 Log-Gabor filters	*83.98*	*92.07*	*86.23*
12MFCCs +12 log-Gabor filters	*91.68*	*94.55*	*92.82*

Table 3 presented results obtained with feature combinations. Reference [21] shows that adding spectral features can improve the classification performance. Thus, we added MFCCs to log-Gabor filters.

In our previous work [22-20], we have shown that the concatenation of 12 log-Gabor filters is achieved the best classification rate compared to using a single filter and the three patches spectrogram with the concatenation of 12 log-Gabor filters. This justifies the use of the concatenation of 12 log-Gabor filters in addition to 12 MFCCs.

The results for the combination of 12 MFCCs and 12 log-Gabor filters are evaluated by the M-SVM-based classifier and HMM-based classifier.

As shown in Fig, we compare the overall classification rate using log-Gabor filters, MFCC and their combination for 10 classes of environmental sounds.

We notice that MFCC features obtain better results than log-Gabor features in four of the examined classes while performing poor results in the case of six other classes; like Door slams (Ds), Dog barks (Db), Gunshots (Gs), Human screams (Hs), Machines (Mc) and Cymbals (Cy).

Log-Gabor filters features achieve better overall, with the exception of two classes (Explosions (Ep) and Glass breaking (Gb) they have the lowest classification rate at 62.50%.

It exist some example in particular the explosion and gunshots classes, which are very similar and contains higher frequencies. According Tab. we note that MFCCs obtain the classification rate 83.87% of this category, log-Gabor filters features were able to yield a classification of rate of 92.07%.

In order to better characterize these sounds, it is preferable to use narrow spectral peaks. MFCC is insufficient to encode narrow-band structure, but log Gabor filters features are effective in doing so.

By adding together Log-Gabor filters and MFCC features, we were able to reach an averaged accuracy rate of order 94.55% in discriminating ten classes.

Besides, there are eight classes that have a recognition rate higher than 90%. We notice that MFCC and Log-Gabor filters features complement each other to obtain the best overall performance.

For classification, we used SVM multi-class: one-versus one.

Table 3 Recognition Rates for Various Features Applied to 1-vs-1 SVMs-Based Classifier

Classes	*Features*		
	MFCCs%	*12Log Gabor filters %*	*MFCC+12 Log-Gabor filters %*
Ds	*75.78*	99.35	*99.76*
Ep	*86.45*	62.50	*88.66*
Gb	*88.63*	78.57	*92.37*
Db	*84.56*	87.50	*90.68*
Pr	*88.94*	83.33	*91.87*
Cv	*88.64*	87.50	*93.38*
Gs	*76.58*	98.21	*99.35*
Hs	*85.36*	94.11	*96.75*
Mc	*79.88*	89.28	*94.83*
Cy	*83.89*	95.83	*97.85*

We can note that the information of MFCCs coefficients is very efficient and suitable to be added to log-Gabor filters. Our experiments confirm this conclusion.

As shown in [27], the fundamental frequency may be similar for different classes for environmental sounds; for this raison and the low dimension of the tested temporal features (ZCR and the average energy) and the frequency features (SRF and SC), these features fail to represent data information. This justifies the use of MFCCs.

The results presented in Table 3 show that log-Gabor filters features are not able to discriminate between classes successfully when used alone like Explosions (Ep) Glass breaking (Gb).The combination of MFCCs and log-Gabor filters separate some classes very well.

Generally, combinations including spectro -temporel domain are useful, because they combine information of the two complementary domains.

MFCCs are spectral features, they characterize the frequency contents. Nerveless, log-Gabor filters features provide temporal and spectral information and also are mostly informative for high frequencies. This justifies the use of 12 MFCCs in addition to 12 log- Gabor filters. As can be shown in Table 3, this combination improves the discrimination ability.

Using One -Against-One SVMs based classifier provides high classification accuracy for the feature combinations.

Moreover, the most informative feature combinations have a large dimension that does not allow the use of GMMs approach, while SVMs are less sensitive to the dimension of the data space.

3.4 Comparison of state-of-the-art methods

Our experimental result was compared to the state-of-the-art methods results.

By comparison with classic descriptors of environmental sounds system already established, we find that the proposed features which based on combination of 12 log-Gabor filters and 12 MFCCs is positioned in the first ranks (92.82%).

Indeed, as illustrated in Table 4 , the combination between 13 MFCCs, 1 RASTA-PLP, 5 Amplitude Descriptor (AD), 1 Spectral Flux (SF), 1 Loudness [39] has given an average classification rate of the order 88.2%.

In [7], the classification system used as features 16 MFCCs+ energy+$\Delta + \Delta\Delta$, achieves a recognition rate is of the order 89.3%.

Moreover, the system of Chu et al. [21] provided a combination of matching pursuit (MP) and MFCCs features. The obtained averaged classification rate is of order 83.9 %.

Other work [27] used a combination between MFCCs, energy, Log energy, SRF (SpectralRoll-Off-Point and SC (Spectral Centroide). The averaged classification rate is of the order 90.64%.

The comparison with these works proves the advantageous of combining the MFCCs and the 12 log-Gabor filters for environmental sound recognition.

Experimental results show that our features are efficient and suitable in spite of their limited number. This can be partly explained by the fact that the spectro-temporal features have the advantage to combine two complementary domains spectral and temporal.

Table 4 Comparison of state-of-the-art methods

Features	Classification Rate(%)
13 MFCCs , 1 RASTA-PLP, 5 Amplitude Descriptor (AD), 1 Spectral Flux (SF), 1 Loudness [39]	88.20
16 MFCCs+energy+$\Delta + \Delta\Delta$ [7]	89.30
Matching Pursuit (MP) + MFCCs [21]	83.14
MFCCs+energy+Log energy+SRF(SpectralRoll-Off-Point+SC(Spectral Centroide) [27]	90.64
Adopted Descriptors using 12 log-Gabor filters+ 12MFCCs	92.82

5 Conclusion

The paper provides a feature extraction method that uses log-Gabor filters to choose a set of spectro-temporel features, which is efficient and physically interpretable.

Log-Gabor filters features can classify sounds where time and frequency features, are not able to capture discriminative properties of the sounds, features of high complexity, such as spectro-temporal coefficients, are well suitable for the environmental sounds classification.

Our experiments proved the advantages of the log-Gabor filters and MFCCs combination in environmental sound classification. The combination with MFCCs ensures more discrimination performance.

The use of SVMs provides a robust system in high dimensions. They are well based mathematically to get good generalization while retaining high classification accuracy.

Using spectro-temporel features as well as the supervised classification method (SVM) gives the best discrimination between specific sound classes.

Compliance with Ethical Standards

Conflict of Interest: The authors declare that they have no conflict of interest.

My Research doesn't involve any Human Participants and/or Animals.

References

[1] V. Peltonen, J. Tuomi, A. Klapuri, J. Huopaniemi, and T. Sorsa, "Computational audiroty scene recognition," presented at the IEEE Int. Conf.Acoustics, Speech Signal Processing, FL, May 2002.

[2] M. Vacher, D. Istrate, L. Besacier, J. F. Serignat, and E. Castelli, "Sound detection and classification for medical telesurvey," in *Proc. IASTED Biomedical Conf.*, Innsbruck, Autriche, Feb. 2004, pp. 395–399.

[3] A Dufaux, L Besacier, M Ansorge, and F Pellandini, Automatic Sound Detection and Recognition For Noisy Environment. In Proceedings of European Signal Processing Conference (EUSIPCO), 1033-1036, (2000).

[4] A Fleury, N Noury, M Vacher, H Glasson and J.F Serignat, Sound and speech detection and classification in a Health Smart Home. 30th IEEE Engineering in Medicine and Biology Society (EMBS), 4644-4647(2008).

[5] D Mitrovic, M Zeppelzauer, H Eidenberger, Analysis of the Data Quality of Audio Descriptions of Environmental Sounds. Journal of Digital Information Management (JDIM), **5**(2), 48-54 (2007).

[6] K El-Maleh, A Samouelian, and P Kabal, Frame-level noise classification in mobile environments. In Proc. ICASSP, 237–240, (1999).

[7] D Istrate, Détection et reconnaissance des sons pour la surveillance médicale. PhD thesis, INPG, France, 2003.

[8] A. Bregman, *Auditory Scene Analysis*. Cambridge, MA: MIT Press, 1990.

[9] M. P. Cooke, *Modeling Auditory Processing and Organisation*. Cambridge, U.K.: Cambridge University Press, 1993.

[10] L. He, M. Lech, N. Maddage, N. Allen, "Stress and Emotion Recognition Using Log-Gabor Filter", Affective Computing and Intelligent Interaction and Workshops, ACII, 3rd International Conference on, 2009, pp.1-6.

[11] L. He, M. Lech, N. C. Maddage and N Allen, "Stress Detection Using Speech Spectrograms and Sigma-pi Neuron Units", Int. Conf. on Natural Computation, 2009, pp.260-264.

[12] M. Kleinschmidt, "Methods for capturing spectro-temporal modulations in automatic speech recognition", Electrical and Electronic Engineering Acoustics, Speech and Signal Processing Papers, Acta Acustica, Vol.88, No.3, 2002, pp. 416-422.

[13] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. Transactions on Computers. vol.42, no.3, pp.300-311, 1993.

[14] L. Wiskott and C. von der Malsburg. Recognizing faces by dynamic link matching. In Axel Wismüller and Dominik R. Dersch, editors, Symposionüber biologische Informations verarbeitung und NeuronaleNetze- SINN '95, pp. 6368, München, 1996.

[15] O. Ayinde and Y.H. Yang. Face recognition approach based on rank correlation of Gabor-filtered images. Pattern Recognition, Vol. 35, no. 6, pp: 1275-1289, June 2002.

[16] C. J. Lee and S. D. Wang. Fingerprint feature extraction using Gabor filters. Electronics Letters, 1999.

[17] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. Pattern Recogn., vol. 24, no. 12, pp.1167-1186, 1991.

[18] J. Daugman. How iris recognition works. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, no.1, pp. 21-30, Jan. 2004.

[19] M. Zhou and H. Wei. Face verification using gabor wavelets and adaboost. In ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition, p. 404-407, Washington, DC, USA, IEEE Computer Society, 2006.

[20] S. Souli, Z. Lachiri, Multiclass Support Vector Machines for Environmental Sounds Classification in visual domain based on Log-Gabor Filters, International Journal of Speech Technology (IJST), vol.16, no.2, pp.203-213, Springer Link, 2013.

[21] S Chu, S Narayanan, and C.C.J Kuo, Environmental Sound Recognition with Time-Frequency Audio Features. IEEE Trans. on Speech, Audio, and Language Processing. **17**(6), 1142-1158, (2009).

[22] S. Souli, Z. Lachiri, On the Use of Time–Frequency Reassignment and SVM-based classifier for Audio Surveillance Applications, International Journal of Image, Graphics and Signal Processing (IJIGSP), vol6, n°12, 2014.

[23] Dennis, J. and Tran, H.D. and Li, H. (2011). Spectrogram Image Feature for Sound Event Classification in Mismatched Conditions. *Signal Processing Letters, IEEE*, 18: 130-133.

[24] F. Auger and P. Flandrin, "Improving the Readability of Time-Frequency and Time- Scale Representations by the Reassignment Method", IEEE Trans. Signal Proc.,Vol.40, No.5,1995 pp.1068-1089.

[25] Kleinschmidt, M. (2002). Methods for capturing spectro-temporal modulations in automatic speech recognition. Electrical and Electronic Engineering Acoustics, Speech and Signal Processing Papers, Acta Acustica, 88:416-422.

[26] Kleinschmidt, M. (2003) .Localized spectro-temporal features for auto-matic speech recognition. In Proc. Eurospeech, pp. 2573-2576.

[27] Rabaoui, A. Davy, M. Rossignol, S. and Ellouze, N. (2008). Using One-Class SVMs and Wavelets for Audio Surveillance. IEEE Transactions on Information Forensics And Security. 3: 763-775.

[28] V. Espinosa-Duro, M. Faundez-Zanuy. Face Identification by Means of a Neural Net Classifier. Proceedings of IEEE 33rd Annual, International Carnahan Conf. on Security Technology, pp. 182-186, 1999.

[29] S.M. Lajevardi, M. Lech. Facial Expression Recognition Using a Bank of Neural Networks and logarithmic Gabor Filters. DICTA08, Canberra, Australia, 2008.

[30] D.J. Field. Statistics of natural, Relations between the images and the response properties of cortical cells. Jour. of the Optical Society of America, pp. 23792394,1987.

[31] Leonardo Software website. [Online]. Available: http ://www.leonardosoft.com. Santa Monica, CA 90401.

[32] Real World Computing Paternship, Cd-sound scene database in real acoustical environments, 2000, http://tosa.mri.co.jp/ sounddb/indexe.htm.

[33] Christopher M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 2003.

[34]V Vladimir, and N Vapnik, An Overview of Statistical Learning Theory. IEEE Transactions on Neural Networks, 10(5), 988-999, (1999).

[35] V Vapnik, and O Chapelle, Bounds on Error Expectation for Support Vector Machines. Journal Neural Computation, MIT Press Cambridge, MA, USA, 12(9), 2013-2036, (2000).

[36] B Scholkopf, and A Smola, Learning with Kernels, (MIT Press, 2001).

[37] C.-W Hsu, C.-J Lin, A comparison of methods for multi-class support vector machines. J. IEEE Transactions on Neural Networks, 13(2), 415-425, (2002).

[38] Lawrence Rabiner and Biing-Hwang Juang, Fundamentals of Speech Recognition, Prentice-Hall, 1993.

[39] D. Mitrovic, M. Zeppelzauer, H. Eidenberger, “Towards an Optimal Feature Set for Environmental Sound Recognition”, Technical Report TR-188-2, 2006.