Using Three Reassigned Spectrogram Patches and Log-Gabor Filter for Audio Surveillance Application

Sameh Souli[1], Zied Lachiri[2], and Alexander Kuznietsov[3]

[1] École Nationale des Ingénieurs de Tunis
Unité de Recherche Signal, Image et Reconnaissance de Formes
BP. 37 Le Belvédère, 1002 Tunis, Tunisie
soulisameh@yahoo.fr
[2] Institut National des Sciences Appliquées et de Technologie
Département Instrumentation et Mesures
INSAT BP. 676 Centre urbain cedex, Tunis, Tunisie
ziedlachiri@enit.rnu.tn
[3] University of applied Sciences Mittelhessen, Wilhelm-Leuschner-Straße, 13
61169 Friedberg, Germany
alexanderkuznietsov@iem.thm.de

Abstract. In this paper, we propose a robust environmental sound spectrogram classification approach; its purpose is surveillance and security applications based on the reassignment method and log-Gabor filters. Besides, the reassignment method is applied to the spectrogram to improve the readability of the time-frequency representation, and to assure a better localization of the signal components. In this approach the reassigned spectrogram is passed through a bank of 12 log-Gabor filter concatenation applied to three spectrogram patches, and the outputs are averaged and underwent an optimal feature selection procedure based on a mutual information criterion. The proposed method is tested on a large database consists of 1000 environmental sounds belonging to ten classes. The averaged recognition accuracy is of order 90.87% which obtained using the multiclass support vector machines (SVM's).

Keywords: Environmental sounds, Log-Gabor-Filter, Mutual Information, Reassignment Method, SVM Multiclass.

1 Introduction

The environmental sounds domain is vast; it includes the sounds generated in domestic, business, and outdoor environments and can offer many services, for instance surveillance and security applications. Recently, some efforts have been interested in detecting and classifying environmental sounds [1], [2]. In the literature, the majority of studies present approaches for classifying sounds using such as acoustic, cepstral, or spectral descriptors. These descriptors can be used as a combination of some, or even all, of these 1-D audio features together [1]. Recently, some efforts emerge in the new research direction, which demonstrate that image

J. Ruiz-Shulcloper and G. Sanniti di Baja (Eds.): CIARP 2013, Part I, LNCS 8258, pp. 527–534, 2013.

processing techniques can be applied in musical [3], and environmental sounds [4]. In our previous work [4], we have showed that spectrograms can be used as texture images. In order to enhance this work, this paper develops method, based on spectrogram reassignment and spectro-temporal components. However, the spectrogram reassignment is an approach for refocusing the spectrogram by mapping the data to time-frequency coordinates that are nearer to the true region of the analyzed signal support [5].Besides, the reassignment method is applied to the spectrogram to improve the readability of the time-frequency representation, and to assure a better localization of the signal components. Indeed, many studies [6] and [7] show that spectro-temporal modulations play an important role in automatic speech recognition (ASR), in particular log-Gabor filters. Our method begins by spectrogram reassignment of environmental sounds, which then was passed through an averaged 12 log-Gabor filters concatenation applied to three spectrogram patches, and finally passed through an optimal feature procedure based on mutual information. In classification step, we use the SVM's with multiclass approach: One-Against-One. This paper is organized as follows. Section 2 describes environmental sound classification system. Classification results are given in Section 3. Finally conclusions are presented in Section 4.

2 Environmental Sound Classification Based on Reassignment Method and Log-Gabor Filters

2.1 Feature Extraction Method

The method consists in using the reassigned spectrogram patch. The aim is to find the suitable part of spectrogram, where the efficient structure concentrates, which gives a better result. We tested our method using log-Gabor filter for three spectrogram patches. We tested for patch number $N_p = 2,3,4,5$, we remark that the satisfactory result is obtained for $N_p = 3$. The idea is to extract three patches from each reassigned spectrogram. The first patch included frequencies from 0.01Hz to 128Hz, the second patch, from 128Hz to 256Hz, and the third patch, from 256Hz to 512Hz. Indeed, each patch goes through 12 log-Gabor filters $\{G_{11}, G_{12}, \ldots, G_{16}, G_{21}, \ldots, G_{25}, G_{26}\}$, followed

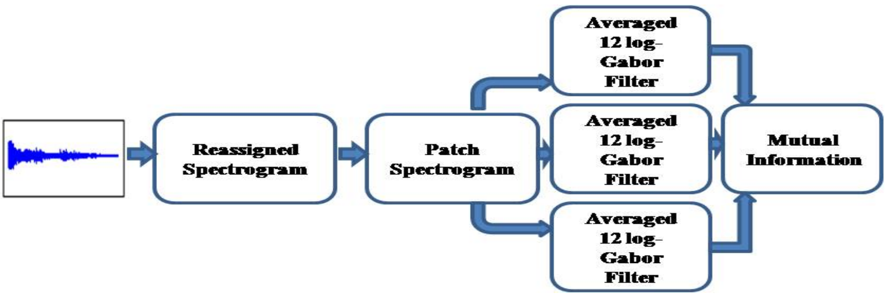

Fig. 1. Feature extraction using 3 spectrogram patches with 12 log-Gabor filters

by an average operation and then, MI feature selection algorithm is used, which constitutes the parameter vector for the classification (Fig.1.).

2.2 Reassignment Method

The spectrogram is the square modulus of the Short Time Fourier Transform $STFT_h(x;t,\omega)$

$$S_h(x;t,\omega) = |STFT_h(x;t,\omega)|^2 \tag{1}$$

$$STFT_h(x;t,\omega) = \int_{-\infty}^{+\infty} x(u)h^*(t-u)e^{-j\omega u}du \tag{2}$$

The disadvantage is manifested by its unseparable kernel allowing the spreads of the time and frequency smoothings bound, and even opposed [8], which leads to the spectrogram a loss of resolution and contrast [9].Hence, the reassignment is going to re-focus the energy spread by the smoothing [10].However, the reassignment application in time–frequency representation provides to run counter to its poor time-frequency concentration.

The reassigned spectrogram obtained by the Short Time Fourier transform (STFT) enhances the concentration of the components in comparison to the spectrogram, and it does not contain any cross terms. The values of the new position of energy contributions $(\hat{t}(x;t,\omega), \widehat{\omega}(x;t,\omega))$ are given by the center of gravity of the signal energy located in a bounded domain centered on (t,ω). These coordinates are defined by the smoothing kernel $\phi_{\mathrm{TF}}(u,\Omega)$ and computed by means of short-time Fourier transforms in the following way [8]:

$$\hat{t}(x;t,\omega) == \mathrm{t} - \mathcal{R}\left\{\frac{STFT_{Th}(x;t,\omega).STFT^*{}_h(x;t,\omega)}{|STFT_h(x;t,\omega)|^2}\right\} \tag{3}$$

$$\widehat{\omega}(x;t,\omega) == \omega + Im\left\{\frac{STFT_{Th}(x;t,\omega).STFT^*{}_h(x;t,\omega)}{|STFT_h(x;t,\omega)|^2}\right\} \tag{4}$$

For more explication, you can see Appendix of [8]. The corresponding equation to the reassignment operators is writing in the following way:

$$\mathrm{MS_h} = \iint \mathrm{S_h}\,(x;\mathrm{t},\omega)\delta\big(\mathrm{t}' - \hat{\mathrm{t}}(x;\mathrm{t},\omega)\big).\,\delta\big(\omega' - \widehat{\omega}(x;\mathrm{t},\omega)\big)\mathrm{dt}\frac{\mathrm{d}\omega}{2\pi} \tag{5}$$

where $\delta(t)$ is the Dirac impulse

We adopted in this work the reassignment method in order to obtain a clear and easily interpreted spectrogram, whose purpose is to improve the classification system performance realized in previous work [11].

2.3 Log-Gabor Filters

Gabor filters have many useful and important properties, in particular the capacity to decompose an image into its underlying dominant spectro-temporal components [6]. The log-Gabor function in the frequency domain can be described by the transfer function $\boldsymbol{G}(\boldsymbol{r},\boldsymbol{\theta})$with polar coordinates [7]:

$$G(r,\theta) = G_{radial}(r).\,G_{angular}(r) \tag{6}$$

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:

$$Re(S(x,y))_{m,n} = s(x,y) * Re\big(G(r_m,\theta_n)\big) \tag{7}$$

$$Im(S(x,y))_{m,n} = s(x,y) * Im\big(G(r_m,\theta_n)\big) \tag{8}$$

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

$$|S(x,y)| = \sqrt{\left(Re\big(S(x,y)\big)_{m,n}\right)^2 + Im(S(x,y))_{m,n}} \tag{9}$$

2.4 Averaging Log-Gabor Filters

The averaged operation was calculated for each 12 log-Gabor filter appropriate for each three reassigned spectrogram patches. The purpose being to obtain a single output array [7]:

$$\left|\hat{S}(x,y)\right| = \frac{1}{N_r N_\theta}\sum_{\substack{m=1\\ n=1}}^{N_r,N_\theta} |S(x,y)|_{m,n} \tag{10}$$

2.5 Mutual Information

The feature vectors were reduced using the mutual information feature selection algorithm. The information found commonly in two random variables is defined as the mutual information between two variables X and Y, and it is given as [12]:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \tag{11}$$

Where $p(x) = Pr(X = x)$ is the marginal probability density function and $p(x) = Pr(X = x)$, and $p(x,y) = Pr(X = x, Y = y)$ is the joint probability density function.

2.6 SVM Classification

The SVM's is a tool for creating practical algorithms for estimating multidimensional functions [13]. In the nonlinear case, the idea is to use a kernel function $K(x_i, x_j)$, where $K(x_i, x_j)$ satisfies the Mercer conditions [14]. Here, we used a Gaussian RBF kernel whose formula is:

$$k(x, x') = exp\left[\frac{-\|x-x'\|^2}{2\sigma^2}\right]. \tag{12}$$

Where $\|.\|$ indicates the Euclidean norm in $\Re^d$. We hence adopted one approach of multiclass classification: One-against-One [15]. For classification with SVM we suggest the cross-validation procedure for kernel parameter γ and the constant of regularization C. Indeed, according to [16], this method consists in setting up a grid-search for γ and C. For the implementation of this grid, it is necessary to proceed iteratively, by creating a couple of values γ and C. In this work, we use the following couples C, γ :C=[$2^{(-5)}$, $2^{(-4)}$,…, $2^{(15)}$] et γ=[$2^{(-15)}$, $2^{(-14)}$, …, $2^{(3)}$].

3 Classification Results and Discussion

Our corpus of sounds comes from commercial CDs [17]. We used 10 classes of environmental sounds as shown in Table 1. 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.

Most of the signals are impulsive. We took 2/3 for the training and 1/3 for the test. Reassigned Spectrograms are extracted through Short Time Fourier Transform with the number of frequency points equal to 512, the smoothing Hanning window is used, which divides the signal into segments of length equal to 256 with 192-point overlap. Indeed, the idea consists in application of reassignment method to 3 spectrogram patches, then passed through a log-Gabor filters concatenation, after that an averaged operation is applied, followed by the mutual information criteria for optimization. Results of our approach are shown in Table 2. Besides, we obtained in this approach an averaged accuracy rate of the order 90.87%. This method leads to an increase approximately 4% of averaged recognition compared to the result obtained when we applied the same method but without using reassignment method which we obtained a

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

Table 2. Recognition Rates for averaged outputs of 3 Reassigned Spectrogram Patches With 12 log-Gabor filters applied to one-against-one SVM's based classifier with Gaussian RBF kernel

3 Reassigned Spectrogram Patches with 12 log-Gabor filters concatenation		
Classes	**Parameters Kernel (c, γ)**	**Classif. Rate (%)**
Ds	$(2^{(-5)}, 2^{(-6)})$	94.87
Ep	$(2^{(-4)}, 2^{(-6)})$	88.75
Cb	$(2^{(-5)}, 2^{(2)})$	78.57
Db	$(2^{(1)}, 2^{(3)})$	89.58
Pr	$(2^{(15)}, 2^{(1)})$	93.75
Cv	$(2^{(-1)}, 2^{(-6)})$	85.71
Gs	$(2^{(-4)}, 2^{(2)})$	95.83
Hs	$(2^{(-3)}, 2^{(-4)})$	95.58
Mc	$(2^{(-4)}, 2^{(-6)})$	92.85
Cy	$(2^{(-3)}, 2^{(-7)})$	93.30

classification rate of the order 86.78 % [11]. The studies of Chu et al. [1] proposed an approach based on combination of matching pursuit (MP) and MFCCs features. This combination gives the average classification rate of 83.9 % in discriminating fourteen classes with GMM classifier. When comparing this result with our system classification rate, we remark that our system has more significant and better results. We notice that our individual features are significantly better in spite of a limited number of features.

To conclude, we compared also our obtained results with the results attained by Rabaoui et al. [2], who used a combination between energy, Log energy and MFCCs features. This research gives an average classification rate of (90.23%) in the environmental sounds classes. It is slightly lower than our proposed method result (90.87%). Moreover, applying the reassignment method on the environmental sound spectrogram enhances the performance of used system. The experimental results reported in this work show that the reassignment method provides a higher improvement in the environmental sounds classification. Therefore, with the reassignment method we can easily interpret the spectrogram signature. In addition, the important point of the reassignment method is the proper choice of smoothing kernel in order to produce simultaneously a high concentration of the signal components [8].The purpose of reassignment method is to build a readable time-frequency representation process. Previous studies [10], [18] show that using reassignment method can improve the detection, the additive sound modeling, and the classification performance. Nevertheless, features extracted from reassigned spectrogram improve the classification results as shown in Table II. SVMs have proven to be robust in high dimensions. Also SVMs are well founded mathematically to reach good generalization while keeping high classification accuracy. The performance of the proposed classification system has been evaluated and compared with our previous work by using a set of synthetic test signals. However, the proposed method maintains overall good performance. The experiments results are satisfactory, which encourages us to investigate better in the reassignment method.

4 Conclusion

In this paper, we propose a robust method for environmental sound classification, based on reassignment method and log-Gabor filters. We show how this method is efficient to classify the environmental sounds. Besides, our method uses an averaged 12 log-Gabor filters concatenation applied to 3 reassigned spectrogram patches. Our classification system obtains good averaged classification result of the order 90.87%.

Furthermore, reassignment method improves classification results. It used as the key element of obtaining an optimal classification compared to our previous methods [11]. In addition, this paper deals with robust features used with one-against-one SVM-based classifier in order to have a system that quietly works, independent of recording conditions. Future research directions will include other methods extracted from image processing to apply in environmental sounds classification and will can be improved while digging deeply into reassignment methods.

References

1. Chu, S., Narayanan, S., Kuo, C.C.J.: Environmental Sound Recognition with Time-Frequency Audio Features. IEEE Trans. on Speech, Audio, and Language Processing 17(6), 1142–1158 (2009)
2. Rabaoui, A., Davy, M., Rossignol, S., Ellouze, N.: Using One-Class SVMs and Wavelets for Audio Surveillance. IEEE Transactions on Information Forensics and Security 3(4), 763–775 (2008)
3. Yu, G., Slotine, J.J.: Fast Wavelet-based Visual Classification. In: Proc. IEEE International Conference on Pattern Recognition, ICPR, Tampa, pp. 1–5 (2008)

4. Souli, S., Lachiri, Z.: Environmental Sounds Classification Based on Visual Features. In: San Martin, C., Kim, S.-W. (eds.) CIARP 2011. LNCS, vol. 7042, pp. 459–466. Springer, Heidelberg (2011)
5. Kelly Fitz, R., Sean Fulop, A.: A unified theory of time-frequency reassignment. Computing Research Repository-CORR, abs/0903.3 (2009)
6. Kleinschmidt, M.: 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 (2002)
7. He, L., Lech, M., Maddage, N., Allen, N.: Stress and Emotion Recognition Using Log-Gabor Filter. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII, Amsterdam, pp. 1–6 (2009)
8. Auger, F., Flandrin, P.: Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method. IEEE Trans. Signal Proc. 40, 1068–1089 (1995)
9. Chassande-Mottin, E.: Méthodes de réallocation dans le plan temps-fréquence pour l'analyse et le traitement de signaux non stationnaires. PhD thesis, Cergy-Pontoise University (1998)
10. Millioz, F., Martin, N.: Réallocation du spectrogramme pour la détection de frontières de motifs temps-fréquence. In: Colloque GRETSI, pp. 11–14 (2007)
11. Souli, S., Lachiri, Z.: Multiclass Support Vector Machines for Environmental Sounds Classification in visual domain based on Log-Gabor Filters. International Journal of Speech Technology (IJST) 16(2), 203–213 (2013)
12. Kwak, N., Choi, C.: Input Feature Selection for Classification Problems. IEEE Trans. on Neural Networks 13, 143–159 (2002)
13. Vladimir, V., Vapnik, N.: An Overview of Statistical Learning Theory. IEEE Transactions on Neural Networks 10, 988–999 (1999)
14. Vapnik, V., Chapelle, O.: Bounds on Error Expectation for Support Vector Machines. Journal Neural Computation 12, 2013–2036 (2000)
15. Hsu, C.-W., Lin, C.-J.: A comparison of methods for multi-class support vector machines. J. IEEE Transactions on Neural Networks 13, 415–425 (2002)
16. Hsu, C.-W., Chang, C.-C., Lin, C.-J.: A practical Guide to Support Vector Classification. Department of Computer Science and Information Engineering National Taiwan University, Taipei, Taiwan (2009)
17. The Leonardo Software website, Santa Monica, CA 90401, http://www.leonardosoft.com
18. Fitz, K., Haken, L.: On the Use of Time-Frequency Reassignment in Additive Sound Modeling. J. Audio Eng. Soc. (AES) 50, 879–893 (2002)