International Conference on Control, Engineering & Information Technology (CEIT'13)
Proceedings Engineering & Technology - Vol.2, pp. 105-109, 2013
Copyright - IPCO

Multiclass Support Vector Machines for Environmental Sounds Recognition with Reassignment Method and Log-Gabor Filters

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

[#]*Signal, Image and pattern recognition research unit*
Dept. of Genie Electrique, ENIT
BP 37, 1002, Le Belvédère, Tunisia
[1]soulisameh@yahoo.fr
[2]ziedlachiri@enit.rnu.tn
[*]*University of applied Sciences Mittelhessen, Wilhelm-Leuschner-Straße,*
13 61169 Friedberg, Germany
[3]alexanderkuznietsov@iem.thm.de

Abstract— **We present a robust environmental sound classification approach, based on reassignment method and log-Gabor filters. 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 database of 10 environmental sound classes. The evaluation system is realized by using the multiclass support vector machines (SVM's) that gave rise to a recognition rate of the order 90.87%.**

Keywords—**Environmental Sound, Log-Gabor-Filter, Mutual Information, Reassignment Method, SVM Multiclass.**

I. 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 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.

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

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

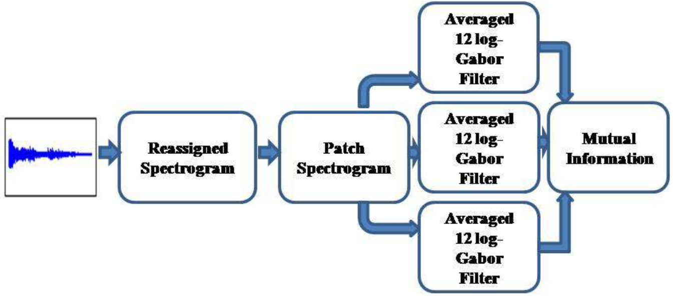
Reassigned Spectrogram
Patch Spectrogram
Averaged 12 log-Gabor Filter
Averaged 12 log-Gabor Filter
Averaged 12 log-Gabor Filter
Mutual Information

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

B. Reassignment Method

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

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

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

Nevertheless, this representation has certain disadvantages. This 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.

In this case the smoothing kernel $\phi_{\mathrm{TF}}(u,\Omega)$ is the Wigner-Ville distribution of some unit energy analysis window $h(t)$, with $\phi_{\mathrm{TF}}(u,\Omega) = \mathrm{WV}(h;u,\Omega)$.

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]:

$$\hat{t}(x;t,\omega) = t - \frac{\int\int u.\mathrm{WV}(h;u,\Omega)\mathrm{WV}(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}}{\int\int \mathrm{WV}(h;u,\Omega)\mathrm{WV}(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}} \qquad (3)$$

$$= \mathrm{t} - \mathcal{R}\left\{\frac{\iint u.Ri^*(h;u,\Omega)Ri\,(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}}{\iint Ri^*(h;u,\Omega)\,Ri(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}}\right\} \qquad (4)$$

$$= \mathrm{t} - \mathcal{R}\left\{\frac{STFT_{Th}(x;t,\omega).STFT^*{}_h(x;t,\omega)}{|STFT_h(x;t,\omega)|^2}\right\} \qquad (5)$$

$$\widehat{\omega}(x;t,\omega) = \omega - \frac{\int\int \Omega.\mathrm{WV}(h;u,\Omega)\mathrm{WV}(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}}{\int\int \mathrm{WV}(h;u,\Omega)\mathrm{WV}(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}} \qquad (6)$$

$$= \omega - \mathcal{R}\left\{\frac{\iint \Omega.Ri^*(h;u,\Omega)Ri\,(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}}{\iint Ri^*(h;u,\Omega)\,Ri(x;t-u,w-\Omega)du\frac{d\Omega}{2\pi}}\right\} \qquad (7)$$

$$= \omega + Im\left\{\frac{STFT_{Th}(x;t,\omega).STFT^*{}_h(x;t,\omega)}{|STFT_h(x;t,\omega)|^2}\right\} \qquad (8)$$

With $\qquad Ri(x;t,\omega) = x(t).X^*(\omega)e^{-j\omega t}$

where $\delta(t)$ isthe Dirac impulse.

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

$$MS_h = \iint S_h\,(x;t,w)\delta\big(t' - \hat{t}(x;t,w)\big).\delta\big(w' - \widehat{w}(x;t,w)\big)dt\frac{dw}{2\pi} \qquad (9)$$

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].

C. Log-Gabor Filters

Gabor filters offer an excellent simultaneous localization of spatial and frequency information [7]. They 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 $G(r,\theta)$ with polar coordinates [7]:

$$G(r,\theta) = G_{radial}(r).\,G_{angular}(r) \qquad (10)$$

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, $\sigma_r$and$\sigma_\theta$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:

$$Re(S(x,y))_{m,n} = s(x,y) * Re\big(G(r_m,\theta_n)\big) \qquad (11)$$

$$Im(S(x,y))_{m,n} = s(x,y) * Im\big(G(r_m,\theta_n)\big) \qquad (12)$$

(x,y) represents the time and frequency coordinates of a spectrogram, and $m = 1,\dots,N_r = 2$ and $n = 1,\dots,N_\theta = 6$ 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:

$$|S(x,y)| = \sqrt{\left(Re\big(S(x,y)\big)_{m,n}\right)^2 + Im(S(x,y))_{m,n}} \qquad (13)$$

D. Averaging of 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]:

$$\left|\hat{S}(x,y)\right| = \frac{1}{N_r N_\theta}\sum_{\substack{m=1\\n=1}}^{N_r,N_\theta} |S(x,y)|_{m,n} \qquad (14)$$

E. 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]:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)log\,\frac{p(x,y)}{p(x)p(y)} \qquad (15)$$

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.

F. SVM Classification

For the classification, we employ a Support Vector Machine. 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:

$$k(x,x') = exp\left[\frac{-\|x-x'\|^2}{2\sigma^2}\right]. \qquad (16)$$

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

We hence adopted one approach of multiclass classification: One-against-One. This approach consists in creating a binary classification of each possible combination of classes, and the result for$k$ classes is $k(k-1)/2$. The classification is then carried out in accordance with the majority voting scheme [15].

III. Classification Results and Discussion

Our corpus of sounds comes from commercial CDs [16]. 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.

TABLE I
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

Most of the signals are impulsive. We took 2/3 for the training and 1/3 for the test. We suggested for classification the cross-validation procedure [17].

We use the following couples:

C,γ :C=[$2^{(-5)}$, $2^{(-4)}$,…, $2^{(15)}$] et γ=[$2^{(-15)}$, $2^{(-14)}$, …, $2^{(3)}$].

Fig. 2 depicts environmental sounds spectrograms and their reassigned representation. Each class contains sounds with very different temporal or spectral characteristics, levels, duration, and time alignment. For example door slams (Ds) present a wide frequency band but with a short duration.

In addition, for the children voices (Cv), we can distinguish the presence of the privilege frequencies. Concerning phone rings (Pr), we notice that they present fundamental frequencies.

We notice a significant improvement in localization of the reassigned spectrogram in time-frequency domain in comparison to the spectrogram obtained by Short Time Fourier Transform. However, as illustrated in Fig. 2 the best localization amongst the considered representations is obtained with the reassignment method application to spectrogram.

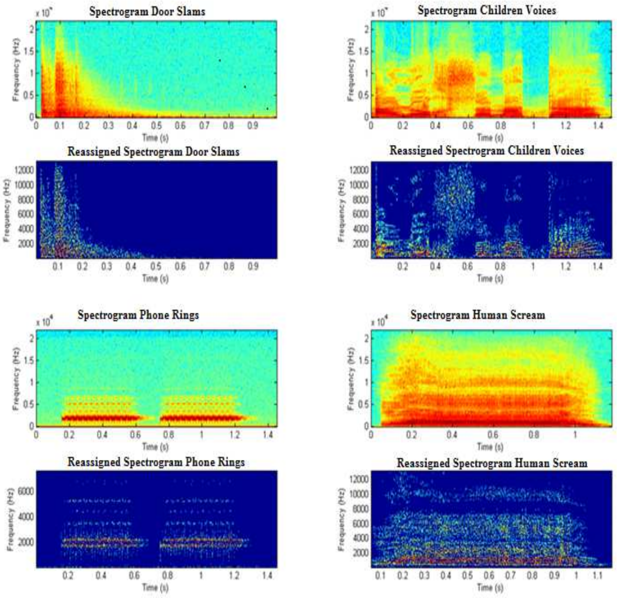

Fig. 2. Spectrograms and their reassigned representation of 4 environmental sound classes

We also remark also that the reassigned spectrogram provides good concentration at lower frequencies, but poor concentration at higher frequencies. In other words, 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 as shown in Fig.2.

The advantage of the STFT in this case is that it does not contain cross terms. Spectrograms are extracted through Short Time Fourier Transform with the number of frequency points equal to 512, the Hanning window is used, which divides the signal into segments of length equal to 256 with 192-point overlap.

In addition, for reassignment spectrogram, we keep the same parameters as used for spectrogram construction. Otherwise, we use the smoothing Hanning window of length 256 with 192-point overlap.

Indeed, the key 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.

TABLE II

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** $(\boldsymbol{c}, \boldsymbol{\gamma})$	**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

Results of third approach are shown in Table 2. Besides, we obtained in this approach an averaged accuracy rate of the order 90.87%.This result is better than the first method result but is slightly lower than the second method result. Also, this method leads to an increase approximately 4% of averaged recognition compared to the result obtained when we applied 12 log-Gabor filters to three patches spectrogram [11] without use of reassignment method.

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.

IV. Conclusions

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] 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*, vol. 17,no. 6 pp. 1142-1158, 2009.

[2] A. Rabaoui, M. Davy, S. Rossignol, and N. Ellouze. "Using One-Class SVMs and Wavelets for Audio Surveillance,"*IEEE Transactions on Information Forensics and Security*. Vol. 3, no.4, pp. 763-775, 2008.

[3] G. Yu, and J. J. Slotine, "Fast Wavelet-based Visual Classification," in Proc. *IEEE International Conference on Pattern Recognition ICPR,* Tampa, 2008, pp.1-5.

[4] S. Souli, Z. Lachiri, "Environmental Sounds Classification Based on Visual Features," *CIARP,Springer,*Chile, vol.7042, pp. 459-466, 2011.

[5] R Kelly Fitz, A Sean Fulop, A unified theory of time-frequency reassignment. *Computing Research Repository- CORR,* abs/0903.3, 2009.

[6] M. Kleinschmidt, "Methods for capturing spectro-temporal modulations in automatic speech recognition,"*Electrical and Electronic EngineeringAcoustics, Speech and Signal Processing Papers, Acta Acustica,* vol.88, pp.416-422, 2002.

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

[8] 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, pp.1068-1089, 1995.

[9] Eric Chassande-Mottin, "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] F Millioz, N Martin, "Réallocation du spectrogramme pour la détection de frontières de motifs temps-fréquence,"*Colloque* GRETSI, 2007, pp.11-14.

[11] 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),* Springer, vol.16, no.2, pp.203-213, 2013.

[12] N. Kwak, C. Choi, "Input Feature Selection for Classification Problems," *IEEE Trans, On Neural Networks*, vol. 13, pp. 143-159, 2002.

[13] V Vladimir, and N Vapnik, "An Overview of Statistical Learning Theory," *IEEE Transactions on Neural Networks,* vol. 10, pp. 988-999, 1999.

[14] V Vapnik, and O Chapelle, "Bounds on Error Expectation for Support Vector Machines,"*Journal Neural Computation, MIT Press Cambridge, MA, USA,*vol12, pp. 2013-2036, 2000.

[15] C.-W Hsu, C.-J Lin, "A comparison of methods for multi-class support vector machines,"*J. IEEE Transactions on Neural Networks*, vol. 13, pp.415-425, 2002.

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

[17] C.-W. Hsu, C-C. Chang, C-J. Lin, "A practical Guide to Support Vector Classification," *Department of Computer Science and Information Engineering National Taiwan University,* Taipei, Taiwan, 2009.

[18] K Fitz and L Haken, "On the Use of Time-Frequency Reassignment in Additive Sound Modeling,"*J.Audio Eng. Soc.(AES),*vol. 50, pp.879-893, 2002.