Proceeding of the IEEE
International Conference on Robotics and Biomimetics (ROBIO)
Shenzhen, China, December 2013

On-line Sound Event Detection and Recognition Based on Adaptive Background Model for Robot Audition

Xinguo Li, Yi Wang, Ting Fan, Dongmang Zhang and Hong Liu*

***Abstract*— It's a natural and convenient way for a robot to interact with outside by robot's ears (i.e. microphones) based on correctly detection and recognition of a sound event. This paper considers sound event detection and recognition in indoor environment where there are varying noises around a robot. To handle the problem of varying background noises, a novel sound event detection and recognition system is developed. Background model update and re-estimation methods are respectively proposed to handle the situations when background noises change slightly or completely. Recognition is then conducted based on the detected sound event by matching it with the noise-corrupted models generated by our proposed combining method modified Parallel Model Combination method (mPMC). mPMC allows modeling the background noise by Gaussian Mixture Model (GMM) of multiple components and can represent the background noise more precisely compared to Single Gaussian Model (SGM). Experimental results show that our adaptive background modeling method attains excellent detection performance in noise-varying conditions and the recognition performance of our proposed mPMC using GMM also outperforms the conventional PMC using SGM in real-world environment with noise varying.**

I. INTRODUCTION

Understanding of a real-world sound signal obtained by robot's ears (i.e. microphones) and behaving correspondingly is a natural and convenient way for robot to interact with the outside. As to the indoor real noisy environments we consider in this paper, there are various sounds/noises around the robot. To be well recognized by robot, sound event detection should be conducted first, since the recognition performance greatly relies on the accurate detection of sound events.

Many researches have been done on sound event detection in the past decades. Dufaux [1] proposed an impulsive sound detection algorithm based on a median filter to analyse the energy variations of the input signal. Vacher [2] presented a detection algorithm based on several wavelet tree mean to detect the beginning of the sound. Shon, Kim and Sung employed Hidden Markov Model (HMM) based hang over scheme to increase speech detection probability and a spectral model-based Voice Activity Detection (VAD) method was also introduced in their study [3]. In the work of [4], Ntalampiras established abnormal and normal sound models to distinguish abnormal events from normal events. Some other works were devoted to detecting a few specific sounds such as gun shots or screams [5][6]. However, few works are focusing on detecting sound events on-line in the situations that background noises change completely, for instance, changing from air condition noise to babble noise.

As to sound event recognition which is conducted after detection, Chu et.al proposed the matching pursuit algorithm to obtain time-frequency features of environment sounds for recognition [7]. A daily sound recognition system was developed in [8] with microphones attached to the environments. A robust environmental sound recognition method was proposed for home automation in [9]. For robot application, Tokutsu et.al. developed a daily sound recognition system using principal component analysis of cepstrum data [10]. However, when facing the practical situations where background noise changes, the conventional recognition methods cannot attain satisfying performance.

In this paper, a novel sound event detection and recognition system is developed to handle the problem of varying background noises. In order to accurately detect the sound events in noise-varying conditions, an adaptive background model is established by updating or re-estimating the current model according to the existing noises. Whether the signal in the current time-window belongs to a sound event is determined by the matching degree of the signal and current background model. Recognition is then conducted based on the detected sound event by matching it with noise-corrupted models, which are generated using our proposed modified Parallel Model Combination method (mPMC) by combining the currently re-estimated background models and clean sound models. PMC was firstly proposed in Varga and Moore's work [11] and refined by Gales and Young in [12][13]. Kim and Hansen proposed a feature compensation method based on Parallel Combined Gaussian Mixture Models (PCGMM) in speech recognition [14]. Our proposed mPMC is extended from Kim and Hansen's work of PMC. The background noises of their work were modeled by single Gussian while Gussian Mixture Models (GMM) with eight components are adopted in our mPMC, which can represent the background more precisely. A novel method that com-

This work is supported by National Natural Science Foundation of China (NSFC, No.60875050, 60675025), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247, 2012AA011705), Scientific and Technical Innovation Commission of Shenzhen Municipality (No.JCYJ20120614152234873, CXC201104210010A, JCYJ20130331144631730, No.JCYJ20130331144716089).

Xinguo Li is with the Shenzhen National Engineering Laboratory of Digital Television Co.,Ltd, Shenzhen, China xgli@neldtv.org

Yi Wang is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School of Peking University, Shenzhen, China wangyi@sz.pku.edu.cn

Ting Fan is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School of Peking University, Shenzhen, China fanting19900126@126.com

Dongmang Zhang is with the Shenzhen National Engineering Laboratory of Digital Television Co.,Ltd, Shenzhen, China dmzhang@neldtv.org

*Corresponding Author Hong Liu is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School of Peking University, Shenzhen, China hongliu@pku.edu.cn

978-1-4799-2744-9/13/$31.00 ©2013 IEEE

bines the background and clean sound models represented by GMM using eight components is also proposed.

The remainder of this paper is organized as follows, a novel sound event detection algorithm based on updating and re-estimating the background model is introduced in Section II. Section III describes the recognition algorithm based on mPMC. In Section IV, experiments and discussions are presented to verify the performance of our proposed method. Conclusions are drawn in Section V.

II. SOUND EVENT DETECTION

In this section, an on-line sound event detection system using an adaptive background model is developed. Considering the varying noises, two situations are taken into consideration to establish the adaptive background model. When background changes slowly and slightly, background model update is conducted while background re-estimation is performed if background has completely changed. The following subsections introduce the background model updating and re-estimating method in detail.

A. Sound event detection algorithm

In this paper, sound event detection proceeds by matching the signal in current time-window with current background model. Background noise is modeled using a GMM with the Gaussian probability-density function as follows:

$$p(x) = \sum_{k=1}^{K} \omega_k N(x|\mu_k, \Sigma_k) \tag{1}$$

K is the number of the Gaussian components. ω_k, μ_k and Σ_k denote the weight, mean vector, covariance matrix of the k_{th} Gaussian component respectively. x is the feature vector and $\sum_{k=1}^{K} \omega_k = 1$.

Whether the signal in the current time window belongs to a sound event is determined by the matching degree between the signal and current background model. On the one hand, to eliminate the influence of some fake sound events such as an impulsive noise, the current signal won't be considered as a sound event unless the matching likelihoods within two consecutive time windows are less than a predefined threshold TH. On the other hand, the precondition of determining current signal as background is that the matching likelihoods within two consecutive time windows are greater than TH, since the stationary part of a sound event may be mistaken for background. This method can reduce the false detection rate and ensure the integrity of the sound event for the following recognition step.

B. Adaptive background model

In order to accurately detect the sound events in noise-varying conditions, an adaptive background model is established by updating or re-estimating the current model according to the existing noises. It is assumed that the preceding seconds of an audio stream is background noise and it is trained as an initial background model. Background model is refreshed adaptively along with the change of noise.

1) Background model update: Background model update is conducted when the background noise changes slightly. If the signal of current time window is determined as background, it will be utilized to update the current background model using Maximum a Posterior (MAP) [15]. Generally, only the mean vector is re-evaluated since it impacts the result primarily. The re-evaluating formula is :

$$\hat{\mu_k} = \frac{\tau_k * \mu_k + \sum_{t=1}^{T} c_{k_t} x_t}{\tau_k + \sum_{t=1}^{T} c_{k_t}} \tag{2}$$

where $c_{k_t} = \frac{\omega_k N(x_t|\mu_k, \Sigma_k)}{\sum_{k=1}^{K} \omega_k N(x_t|\mu_k, \Sigma_k)}$, and x_t is the adaptive data. λ = (ω_k, μ_k, Σ_k) is the k_{th} Gaussian component of the current background model. More details of the formula are in [10]. $\tau_k = 1/(m_k - \mu_k)$, m_k denotes the mean vector of the adaptive data and τ_k controls the dependent degree of the adaptive data to μ_k.

2) Background model re-estimation: The detection method mentioned above, however, will become invalid when the background noise changes completely. More specifically, when a new background noise appears, it will be falsely detected as a sound event since the matching likelihoods between signals and current background models in multiple time windows will be less than TH. This is a difficult situation that conventional sound event detection methods cannot handle well. Distinguishingly, a novel background model re-estimation approach is proposed in our paper according to the situation when background noise changes completely.

Algorithm 1: Background model re-estimation

Input: $x(t)$, Θ
Output: re-estimated Θ

Do
If *Matching*$(\Theta, x(t)) < TH$
  $issound \leftarrow issound + 1$;
Endif
If $issound == LEN$
  For $i = 1 : floor(LEN/2)$
    $isnewbg \leftarrow 1$;
    $\Theta\,(t - LEN + i) \leftarrow$ *Training*$(x(t - LEN + i))$;
    For $j = t - LEN + 1 + i : t$
      If *Matching*$(\Theta\,(t - LEN + i)\,, x(j)) < TH$
        $isnewbg \leftarrow 0$;
        break;
      Endif
    Endfor
    If $isnewbg == 1$;
      $\Theta \leftarrow \Theta\,(t - LEN + i)$;
      $issound \leftarrow 0$;
      break;
    Endif
  Endfor
Endif
Until EOF

Background re-estimation method is illustrated in Algorithm 1. $x(t)$ is the signal within current time window t along

with the increase of t while Θ is the current background model. $issound$ accumulates the number of windows whose corresponding signals are consecutively determined as sound and TH denotes the likelihood threshold. LEN is a fixed number predefined to measure the general length of a sound event we are going to recognize. *Training* is a function of GMM model training and *Matching* calculates the likelihood of current signal within current background model. $isnewbg$ symbolizes whether there is a new background noise.

When signals in consecutive LEN windows are detected as sound, a new kind of background noise is likely to appear. Whether the signals within LEN windows are stationary or not will be checked in order to identify the appearance of a new kind of background. $isnewbg$ will be set to 1 if the signal is determined stationary and a new background model will be re-estimated based on the latest window of the signal among the LEN windows. More details are presented in Algorithm 1.

Fig.1 is the general flow chart of our novel detection method which can handle the situations when background noise changes slightly or completely.

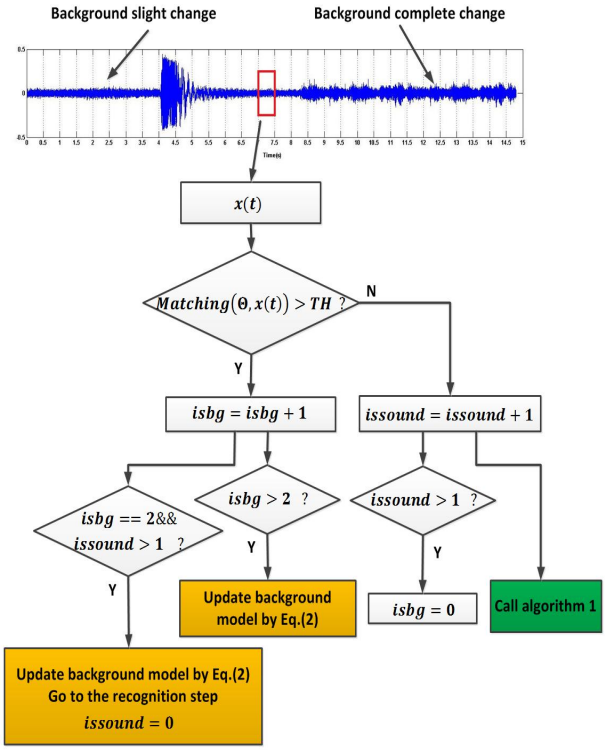

Fig. 1. Flow chart of sound event detection algorithm

III. SOUND EVENT RECOGNITION BASED ON MODIFIED PMC

In this section, a modified Parallel Model Compensation(mPMC) method based on GMM with 8 components is introduced for sound event recognition, taking advantages of the updated or re-estimated background model in real-time above.

Recognition performance in real-world environments with various noises declines significantly due to the difference between testing and training data. It's necessary to choose an approach to re-estimate the noise-corrupted model parameters to obtain satisfying performance. With the availability of clean sound models and background noise models, PMC is a practical model-based compensation method which generates noise-corrupted models by combining clean sound models and background noise models.

PMC assumes that sound and noise are dependent and are additive in the linear-spectral domain. Since GMM parameters based on mel-frequency cepstral coefficients (MFCC) belong to cepstral domain, inverse Discrete Cosine Transformation (DCT) is firstly applied to transform the mean and covariance in cepstral domain to log-spectral domain. Then, the mean and covariance in log-spectral domain are transformed to linear spectral domain. The transformation rules are discussed in [13].

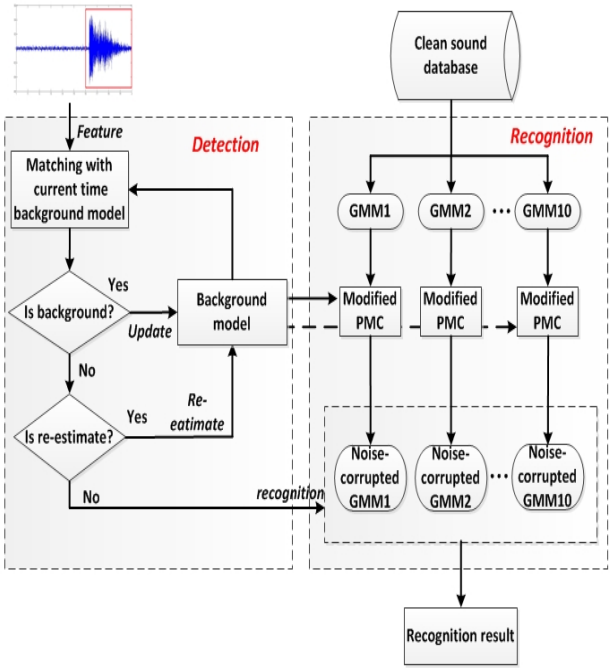

Fig. 2. Diagram of detection and recognition system

In order to represent the background more precisely, GMMs with eight components are adopted to model the adaptive background model, extended from Kim and Hansen's work [14]. The proposed mPMC method explains the rules combining the background GMM models and clean sound models in linear spectral domain.

Firstly, the mean vector and covariance in linear spectral domain can be obtained by the transformation rules mentioned above. Suppose that parameters in linear spectral domain of the clean sound model λ_x and background model λ_n are described as follows:

$$\lambda_x = (\omega_{xk}, \mu_{xk}, \Sigma_{xk}), k = 1, 2, ..., K \quad (3)$$

$$\lambda_n = (\omega_{nk}, \mu_{nk}, \Sigma_{nk}), k = 1, 2, ..., K \quad (4)$$

Then our combining method proceeds as follows:

$$\hat{\mu}_{xk} = g\mu_{xk} + (1-g)\sum_{k=1}^{K}\omega_{nk}\mu_{nk} \tag{5}$$

$$\hat{\Sigma}_{xk} = g^2\Sigma_{xk} + (1-g)^2\sum_{k=1}^{K}\omega_{nk}\Sigma_{nk} \tag{6}$$

$$\hat{\omega}_{xk} = \omega_{xk} \tag{7}$$

here, $\hat{\mu}_{xk}$, $\hat{\Sigma}_{xk}$, $\hat{\omega}_{xk}$ are parameters of the k_{th} Gaussian component of the noise-corrupted model in linear spectral domain. g denotes the gain factor which can be obtained by the formula:

$$g = (E - E_n)/E \tag{8}$$

where $E = mean(x^2(t))$ and $E_n = mean(noise^2(t))$, respectively referring to the average energy of current detected sound signal $x(t)$ and the average energy of noise segment $noise(t)$. $noise(t)$ is the signal in latest several time windows before the time window that detected as current sound event.

After combining the linear parameters of sound event models and background models, mapping from linear spectral domain back to log-spectral domain and back to the cepstral domain are simply the reverse operation. Finally, recognition is conducted using the generated noise-corrupted models on the detected sound event in real-time. The most probable class is determined by the max likelihood between current signal and noise-corrupted models of different sound event we are going to recognize.

The diagram of detection and recognition method of this paper is shown in Fig.2.

IV. EXPERIMENTS AND DISCUSSIONS

Ten kinds of sound events are discussed in our experiment including door slam, glass break, hand clap, knock, pat desk, scream, whistle, laugh, phone ring and printing. Three databases are established in indoor environment. DB.1, eleven audio streams containing 570 sound events in noiseless condition. DB.2, 630 segments of 10 kinds of sound events in noiseless condition. DB.3, an audio stream with 670 sound events in real-world environment with varying background air-conditioner (appear or disappear), computer engine noise and babble (broadcasted by loudspeaker). These signals are digitized at a sampling rate of 11,025*HZ*, and 16 bits per sample.

Four comparison experiments are conducted on the databases above to verify the effectiveness of our proposed sound event detection method as well as the mPMC approach used in the recognition step.

Firstly, a comparison of detection performance using three background model estimation methods is presented. Three model estimation methods are, (1) Model estimation based on initial background without update or re-estimation. (2) Model estimation based on background update only. (3) Model estimation based on both of update and re-estimation. The background noises are modeled by GMMs with eight components based on 13 dimensional normalized MFCC features.

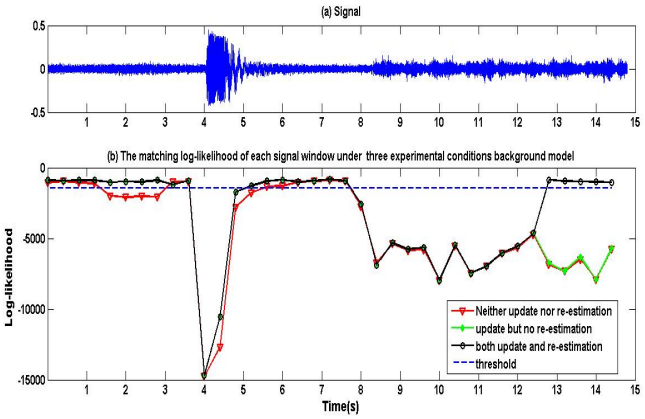

Fig. 3. The matching log-likelihood of each signal window with three background model estimation methods

Fig.3 shows the detection performance of three background model estimation methods experimented on an audio stream that contains a sound event embedded in varying background noise including air-conditioner and babble. The background noises cover both of the situations where noise changes slightly as well as completely.

Fig.3(b) refers to the matching log-likelihood between the signal of current time window and current background model using three estimation methods. The blue dashed line in Fig.3(b) refers to the threshold TH used to distinguish sound event from background noise, which is set empirically.

It can be seen that the sound event can be correctly detected by all of the three methods. However, as to the slightly changing air-conditioner noise in the front, part of the red line (triangle) falls below the threshold line, indicating that background noise is falsely determined as a sound event. This is reasonable since the model estimation of the red line is simply based on the initial background without update or re-estimation, which can not adapt to the slightly varying noise and thus leads to an incorrect determination. When the background noise completely changes into the babble noise, the detection performances of the three methods also differ. The new noise is incorrectly determined as a sound event by both of the corresponding methods in red line (triangle) and green line (rhombus), resulting from the absence of re-estimation. Distinguishingly, according to the completely changing noise, our proposed model re-estimation method can recognize the newly appeared noise as background instead of a sound event. Generally, the proposed adaptive background modeling method outperforms the other two methods since it can effectively distinguish sound events from varying background noises.

Secondly, a comparison of the detection performance of modeling the background by GMM and SGM, as well as the recognition performance of our proposed mPMC using GMM and Kim's work [14] using SGM under real-world environment are presented. Ten kinds of clean sound event models are trained on the database of DB.2 using GMM with eight components and testing experiment is conducted on the

database DB.3 of real-world environment .

Table 1 shows on-line detection and recognition performance in the real-world environment (DB.3) using different methods. "GMM(1)", "GMM(2)" and "GMM(3)" respectively refer to modeling the background by GMM of eight components "without update or re-estimation", "update but no re-estimation" and "both update and re-estimation". "SGM(3)" refers to modeling the background by Single Gaussian Model and performs background model update as well as re-estimation. CDR, FDR and CRR denote Correctly Detection Rate, Falsely Detection Rate and Correctly Recognition Rate respectively

It can be seen that among the three background model estimation methods using GMM of eight components, our proposed "GMM(3)" attains the best detection performance due to the update and re-estimation of background model when background noise changes. Besides, "GMM(3)" also outperforms "SGM(3)" with higher CDR and lower FDR since SGM could not model the background noises as precisely as GMM with eight components, especially for the babble noise.

TABLE I
ON-LINE DETECTION AND RECOGNITION PERFORMANCE OF REAL-WORLD AUDIO STREAM USING DIFFERENT METHODS

Method		Detection		Recognition	
		CDR	FDR	$no-PMC$	$mPMC$
GMM	(1)	81.11%	14.52%	59.64%	83.11%
	(2)	85.06%	9.64%	60.76%	84.56%
	(3)	95.74%	2.38%	62.79%	85.70%
SGM	(3)	94.89%	6.75%	59.37%	85.42%

Better detection performance can result in better recognition performance, which can be reflected in that the mPMC recognition performance of "GMM(3)" and "SGM(3)" with better detection performance outperform "GMM(1)" and "GMM(2)". Besides, the recognition methods using mPMC generally attain better performances than those without PMC due to the compensation between clean models and noisy signals. Comparing the recognition performance of our mPMC with eight components and PMC using SGM in [14], we can also see that mPMC using GMM has restricted improvement. It is reasonable since our proposed mPMC merges eight Gaussian components of background in linear spectral domain together for combination, which is similar with the combining method of PMC using SGM in [14].

In general, our proposed adaptive background modeling method can better detect sound events when background noises change slightly or completely. Additionally, our mPMC using GMM also outperforms PMC using SGM [14] slightly.

Another experiment is conducted to explore the detection and recognition performance of our approach under several conditions with different SNRs (Signal-to-Noise Ratio). Air conditioner and babble noises are added into the 11 clean audio streams (DB.1) with different SNRs ($0dB$, $10dB$, $20dB$). An audio stream with 670 occurrences of sound events (DB.3) in real-world environment with an average SNR of $13.6dB$ is also tested.

TABLE II
ON-LINE DETECTION AND RECOGNITION PERFORMANCE UNDER DIFFERENT NOISY CONDITIONS

Condition		CDR	FDR	CRR
Air-conditioner	$0dB$	78.94%	2.76%	43.54%
	$10dB$	94.76%	2.69%	78.25%
	$20dB$	100%	1.16%	91.26%
Babble	$0dB$	78.64%	2.43%	37.00%
	$10dB$	93.65%	1.65%	73.19%
	$20dB$	96.83%	1.08%	87.49%
Real-world	$13.6dB$	95.74%	2.38%	85.70%

Table 2 shows the on-line detection and recognition performance under conditions with different SNRs. It can be seen that the CDR decreases with SNR declining, which results from the fact that the feature characteristics of sound events are weakened by intense noises. It is assumed that the current signal is considered as sound when the matching likelihoods between the signal and current background model in two consecutive time windows are both less than TH, which aims to ensure the integrity of the signal to be recognized in the next step. Although this assumption may lead to missing detections, a small number of missing detections has little influence on CDR.

Besides, it can be noted that the noise intensity has little influence on FDR since FDR primarily depends on the stability of noises. The last column of Table 2 indicates that the CRR also decreases with the SNR declining. One reasonable explanation is that errors may occur in compensation using mPMC when SNR is low. Another proper reason is that signals in part of time windows that belong to current sound event are falsely determined as background in the existence of intense noise.

Finally, recognition performance of ten kinds of sound events under off-line noiseless condition and on-line real-world environment using our approach is explored. As to the off-line condition, half of the data in DB.2 (630 sound event segments in noiseless condition) is used to train ten GMMs representing ten kinds of sound events in advance and the remaining data is tested off-line. On-line experiment is conducted in DB.3 of real-world environment with the existence of practical noises.

Fig.4 compares the recognition performances of ten kinds of sound events under off-line noiseless condition and on-line real-world environment respectively. The darker bar shows the recognition performance under off-line condition while the lighter bar refers to the on-line condition.

It can be seen that the on-line recognition performance is relatively worse than that of the off-line condition. This can be explained by the reason that sound event segments tested

off-line are recorded in noiseless environment while the on-line condition explores the recognition performance in real-world environment including varying background noises. Besides, errors primarily exist in the recognition step of off-line condition since the testing data is segmented sound events. As to on-line condition, errors may occur both in the sound event detection step and the mPMC recognition step according to experiments on the audio steam in real-world environment.

Looking into the mis-recognized examples, it can be found that laugh is sometimes mis-classified as scream while hand clapping and patting desk are easily confused with each other. Generally, most of sound events are correctly recognized which verifies the effectiveness of our approach.

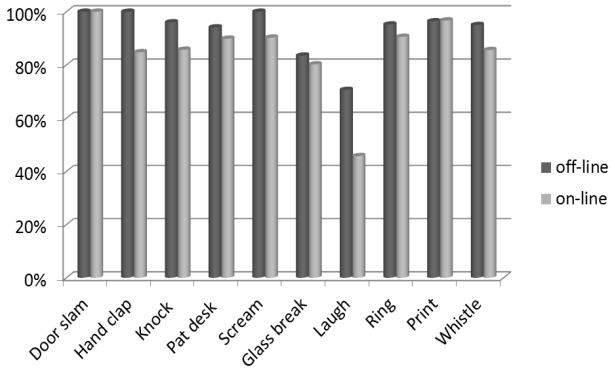

Fig. 4. Recognition performance of 10 kinds of sounds

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a novel sound event detection and recognition system is proposed to handle the problem of varying background noises. Background model update and re-estimation methods are proposed to respectively handle the situations when noise changes slightly and completely. Taking advantage of the current background model refreshed above, recognition is then conducted based on the detected sound event by matching it with noise-corrupted models using our proposed mPMC. Experimental results show that background model update and re-estimation methods attain excellent detection performance in noise-varying conditions and the recognition performance of our proposed mPMC using GMM also outperforms conventional PMC using SGM in real-world environment with noise varying.

Our future works will focus on designing a more effective combining method of clean sound event model and background model so as to obtain greater improvement in real-world environment with noise varying.

REFERENCES

[1] A. Dufaux, L. Besacier, M. Ansorge, and F. Pellandini, "Automatic sound detection and recognition for noisy environment", *in European Signal Processing Conference*, Tamper, Finland, 2000, pp. 1033-1036.

[2] M. Vacher, D. Istrate, and J.-F. Serignat, "Sound detection and classification through transient models using wavelet coefficient trees", *in European Signal Processing Conference*, Vienna, Austria, 2007, pp. 1171-1174.

[3] J. Sohn, N.S. Kim, and W. Sung, "A statistical model-based Voice Activity Detection.", *in IEEE Signal Processing Letters*, vol.6, no.1, 1999, pp. 1-3.

[4] S. Ntalampiras, I. Potamitis, and N. Fakotakis, "Probabilistic novelty detection for acoustic surveillance under real-world conditions", *in IEEE Transaction on Multimedia*, vol.13, no.4, 2011, pp. 713-719.

[5] C. Clavel, T. Ehrette, and G. Richard, "Event detection for an audio-based surveillance system", *in IEEE International Conference on Multimedia and Expo*, Amsterdam, Netherlands, 2005, pp. 1306-1309.

[6] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection and localization for audio-surveillance systems", *in IEEE International Conference on Advanced Video and Signal Based Surveillance*, London, UK, 2007, pp. 21-26.

[7] S. Chu, S. Narayanan, and C-C. Jay Kuo, "Environmental sound recognition with time-frequency audio features", *in IEEE Transaction on Speech, Audio, and Language Processing*, vol.17, no. 6, 2009, pp. 1142-1158.

[8] C. Jianfeng, K. Alvin Harvey and et.al. "Bathroom activity monitoring based on sound". *Pervasive Computing: Lecture notes in Computer Science*, Vol.3468, 2005, pp. 47C61.

[9] J. C. Wang, H. P. Lee, J. F. Wang, and C. B. Lin, "Robust environmental sound recognition for home automation", *in IEEE Transaction on Automation Science and Engineering*, vol.5, no.1, 2008, pp. 25-31.

[10] S. Tokutsu, K. Okada, and M. Inaba. "Discrimination of daily sounds for humanoids understanding situations". *In Proceedings of the 25th annual conference of the Robotics Society of Japan*,, Chiba, Japan, 2007, p. 1H36.

[11] A. P. Varga, R. K. Moore, "Hidden Markov model decomposition of speech and noise", *in IEEE International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, US, 1990, pp. 845-848.

[12] M. J. F. Gales and S. J. Young, "Robust speech recognition in additive and convolutional noise using parallel model combination", *in Computer Speech and Language*, 1995(9), pp. 289-307.

[13] M. J. F. Gales, S. J. Young, "Robust continuous speech recognition using parallel model combination", *in IEEE Transaction on Speech Audio Process*, vol.4, no.5, 1996, pp. 352-359.

[14] W. Kim, J. H. L. Hansen, "Feature compensation in the cepstral domain employing model combination", *in Speech Communication*, vol.51, no.2, 2009, pp. 83-96.

[15] J. L. Gauvain, C. H. Lee, "Maximum a posterior estimation for multivariate Gaussian observations of Markov Chains", *in IEEE Transaction on Speech and Audio Processing*, vol. 2, no. 2, 1994, pp. 291-298.