IMPROVING EFFICIENCY AND RELIABILITY OF GUNSHOT DETECTION SYSTEMS

Talal Ahmed, Momin Uppal and Abubakr Muhammad

Dept of Electrical Engineering, LUMS School of Science and Engineering, Lahore, Pakistan

ABSTRACT

In this paper, we focus on setting up a gunshot detection system with high detection performance, robustness to noise and low computational complexity. To achieve these objectives, we formulate a two-stage approach with a less costly impulsive event detection framework followed by a relatively more complex gunshot recognition stage. To improve detection performance of the gunshot recognition stage, we propose a template matching measure in conjunction with the eighth order linear predictive coding coefficients to train a support vector machine classifier. Using an extensive audio database, we were able to achieve a better gunshot recognition performance than with the well-known existing features used for gunshot detection.

Index Terms— Gunshot detection systems, Acoustic signal processing, Event detection, Support vector machines, Template matching

1. INTRODUCTION

Acoustic gunshot detection systems have been proposed as a tool for the law-enforcement agencies to detect and report gunfire. In addition, such systems also have military applications where they could allow soldiers on the battlefield to not only detect, but also localize enemy fire. Different characteristics of a gunshot audio signal can be exploited for detection of a gunshot acoustic. A gunshot acoustic consists of a muzzle blast that lasts for about three milliseconds, sound of mechanical action associated with operation of a firearm and possibly a shockwave when the bullet is travelling at supersonic speed [1].

A field employable gunshot detection system should have a high detection rate, minimal false alarm rate, robustness to background noise, and low processing time. Unlike some of the past work that is focused on maximization of gunshot detection performance [2][3][4], the focus of our work is on the system-level design of a field employable gunshot detection system that exhibits all the aforementioned qualities. To decrease computational complexity, we propose a multi-level system architecture consisting of an impulsive event detection block followed by a gunshot detection block. The event detection block ensures that the more complex gunshot detection block is activated only in the case of an acoustic event. Thus, because of the high complexity of feature extraction algorithms in the gunshot recognition stage, the objective of the event detection block is to minimize computational load on the succeeding stage. But, the decrease in computational complexity comes at a cost: lower the sensitivity of the event detection stage, lower the computational load on the succeeding gunshot recognition stage but higher the missed event detection rate and vice versa. Thus, a trade-off between computational complexity of the system and missed event detection rate has to be made in the event detection stage. Three impulsive event detection schemes are proposed in [5] that depend on evolution of power sequence obtained from incoming audio signal segments. One of these methods, that uses a noise-adaptive threshold on the power sequence for pulse detection, provides a nice compromise between detection performance and system response delay [5]. Though this method allows tuning of the threshold with sensitivity of the scheme, there is no way to set the threshold to meet any particular missed event detection rate requirement. Modeling the energy of a signal segment as either chi-square or non-central chi-square distributed, we build an analytical framework on top of this impulsive event detection method such that the event detection threshold can be set to precisely meet any missed detection rate requirement while minimizing computational load on the more complex feature extraction stage.

When a particular audio segment is flagged by the event detection block, it is passed to the gunshot recognition stage. The problem of gunshot recognition is particularly difficult due to the countless possibilities of non-gunshot impulsive audio events in the operating environment. The recognition system should be able to distinguish a gunshot from sounds such as that of a clap, door slam, fire-cracker etc. It is important to have the false positive rate as low as possible, since a high false positive rate may lead to lower productivity and lack of confidence for law enforcement personnel in gunshot detection systems [6]. In recent work, an audio event detection system using two parallel Gaussian mixture model (GMM) classifiers has been proposed for the detection of screams

and gunshots [2]. The gunshot detection problem has also been pursued as a maximization task employing a dynamic programming solution to the detection problem [3]. More recently, the use of correlation against a gunshot template has been proposed by Freire and Apolinario [7]. The superior performance of this detection feature was then reported under noisy environments in comparison with other known methods of gunshot detection [4]. However, we have found in our experiments that even though the correlation method of [7] gives a high true positive rate (TPR), it yields an undesirably high false positive rate (FPR). As a means of reducing this FPR, we propose using cross-correlation maximum against a gunshot template in conjunction with eighth order linear predictive coding (LPC) coefficients [8] as features with a gaussian radial basis function (RBF) Kernel for a support vector machine (SVM) classifier [9]. Using an extensive audio database consisting of gunshot recordings, door slams, ticks, and human voice for an eight-fold cross validation experiment, we were able to achieve a 97.62% true positive rate and a negligible false positive rate on our audio database.

Thus, our contributions towards building a field deployable gunshot detection system are two-fold: First, we provide an analytical framework for a segment power based event detection framework that allows us to meet the missed detection rate requirements of the event detection stage while minimizing computational load on the gunshot detection system. Second, we propose the use of SVMs with template correlation maximum, LPC coefficients and the kernel trick to get high gunshot detection rate while minimizing the FPR.

2. ACOUSTIC EVENT DETECTION

The event detection task over an audio segment can be formulated as a binary hypothesis testing problem:

H_0: segment contains noise only

H_1: segment contains an acoustic event

The audio signal segment at time t can be written as

$$r(t) = hs(t) + n(t) \tag{1}$$

where $h = 0$ under hypothesis H_0 and $h = 1$ under hypothesis H_1, $s(t)$ represents the acoustic event and $n(t)$ represents zero-mean additive white gaussian noise process with power spectral density $\frac{N_0}{2}$.

Over a sampling period T, we use the normalized received signal energy given by

$$Z = \frac{2}{N_0}\int_0^T r^2(t)\,\mathrm{d}t \tag{2}$$

as a decision statistic. According to [10], Z follows a chi-square distribution with $2WT$ degrees of freedom under hypothesis H_0, where W is the positive bandwidth of the received signal. Alternatively, under hypothesis H_1, Z follows non-central chi-square distribution with $2WT$ degrees of freedom and a non-centrality parameter 2γ, where γ is the signal-to-noise (SNR) ratio of the received signal [10]. Defining $a = TW$, the probability density function of the decision statistic Z can be given as

$$f_Z(z) = \begin{cases} \frac{1}{2^a\Gamma(a)} z^{a-1} e^{-\frac{z}{2}} & H_0, \\ \frac{1}{2}\left(\frac{z}{2\gamma}\right)^{\frac{u-1}{2}} e^{-\frac{2\gamma+z}{2}} I_{u-1}(\sqrt{2\gamma z}) & H_1. \end{cases} \tag{3}$$

where $\Gamma(.)$ is the gamma function and I_n is the n^{th} order modified Bessel function of the first kind [11]. The probability of missed detection and false alarm are

$$P_m = Pr(Z < \lambda | H_1), \tag{4}$$

$$P_{fa} = Pr(Z > \lambda | H_0), \tag{5}$$

respectively, where λ is the decision threshold on the energy statistic Z. Using (3), we obtain

$$P_m = 1 - Q_a(\sqrt{2\gamma}, \sqrt{\lambda}), \tag{6}$$

$$P_{fa} = 1 - \frac{\gamma(a, \frac{\lambda}{2})}{\Gamma(a)} = \frac{\Gamma(a, \frac{\lambda}{2})}{\Gamma(a)}, \tag{7}$$

where Q_a is the generalized Marcum Q-function [12], $\gamma(.,.)$ is the lower incomplete gamma function and $\Gamma(.,.)$ is the upper incomplete gamma function [11].

The performance of the event detection block can be characterized by P_{fa}; the lower the P_{fa}, the lower the number of false events reported to the gunshot recognition block and lower is the computational load on the more complex feature extraction stage. However, a more critical parameter associated with event detection is the missed detection rate (P_m). Events missed in the event detection stage are lost, whereas falsely reported events can still be removed in the gunshot recognition stage. Our design philosophy for the event detection scheme is to set the decision threshold λ to meet a constant missed detection rate (CMDR) requirement, so that the optimal trade-off between P_m and P_{fa} can be achieved under system constraints on P_m. From (6), it can be observed that λ can be set so that the CMDR requirement is always met, given that the SNR can be estimated (as described next). We design our event detection scheme such that only the events with segment signal energy above a fixed minimum value of $E_{s_{\min}}$ are detected. This value $E_{s_{\min}}$ would depend on the coverage radius of the detection system, minimum sound pressure level at the gunshot source, and gain of the microphone. Noise energy E_n can be estimated using a long-term median filter on energy values of the previous audio signal segments [5]. The length of the median filter can be decreased so that the detection scheme responds to impulsive events only.

Given an E_n estimate, the SNR is estimated on the basis of worst case segment power $E_{s_{\min}}$ under hypothesis H_1. Thus, the decision threshold λ can be set on the basis of a CMDR and $\text{SNR}_{\min} = \frac{E_{s_{\min}}}{E_n}$ by finding numerical solution for λ from (6). Our scheme ensures that a CMDR requirement for the event detection stage is always met, given that segment power remains above $E_{s_{\min}}$ under hypothesis H_1.

3. SVM-BASED GUNSHOT RECOGNITION

Audio segments flagged by the event detection stage are passed onto the gunshot recognition block. In this section, we describe the audio features, the classification method, and the experimental setup we've used to evaluate performance variation of the gunshot recognition system with different feature combinations and kernel types.

Let the flagged segment comprise of N samples, with each sample denoted by $x_f[n]$, $n = 0, \ldots, N-1$. Let $t[n]$ denote the samples of the template, where the template segment is zero-padded to make sure that it also comprises of N samples. The flagged segment, as well as the template are normalized between 1 and -1 before calculating the cross-correlation as

$$R(m) = \begin{cases} \sum_{n=0}^{N-m-1} x_f[n+m]t[n] & m \geq 0, \\ \sum_{n=0}^{N+m-1} x_f[n]t[n-m] & m < 0. \end{cases} \quad (8)$$

The cross-correlation maximum for the flagged segment is found by picking the maximum $R(m)$. Other features extracted from the flagged segment for comparison are the eighth order Linear Predictive Coding (LPC) coefficients and the first 13 Mel-Frequency Cepstral Coefficients (MFCC) [8]. LPC coefficients are calculated using MATLABs Signal Processing Toolbox, and MFCC coefficients are calculated using the auditory toolbox by Internet Research Corporation [13]. The aforementioned features were extracted from our audio database to evaluate gunshot detection performance of the different feature types. Once a database of relevant feature sets corresponding to both gunshot and outsider signals was built, eight-fold cross-validation [14] was used to train and test SVMs. More details about the experimental setup are mentioned in Section 4.

4. EXPERIMENTAL EVALUATIONS

In this section, we describe the specifics of our audio database, the computational gain of the even detection scheme, and the detection performance of the different feature sets introduced in Section 3.

4.1. Data Acquisition

We acquired an audio database of gunshots by making recordings at a local firing range using a standard PC sound card. An extensive database of G3 and MP5 gunshots was acquired for different shooter distances of 100 meters (m), 200 m, and 300 m. All audio signals were sampled at 44.1 kHz with 8-bit quantization. Moreover, sounds like claps, door slams, ticks, and random people talking were recorded to be used as outsider signals for the gunshot recognition system. In all, our acquired audio database amounted to 434 audio clips consisting of 332 gunshot and 102 outsider signals[1].

4.2. Evaluating Event Detection Stage

In this section, we use the analytical results derived in Section 2 to evaluate performance of the event detection stage. The event detection stage is designed such that given a value for $E_{s_{\min}}$, the detection threshold adapts with the noise power estimate via $\text{SNR}_{\min}$ such that a constant missed detection rate is maintained. However, the computational gain of this stage is directly linked with the false alarm rate; the lower the false alarm rate, the lesser the load on the more complex gunshot recognition stage. Using numerical solutions to (7), we plot in Figure 1 the false alarm rate variation of the event detection stage with the system parameter $\text{SNR}_{\min}$ for a given missed detection rate of 5% and an audio segment size of 1300 samples. The figure shows that the false alarm rate drops to about 67% at a $\text{SNR}_{\min}$ of 15 dB, and about 2% at a $\text{SNR}_{\min}$ of 20 dB, which is not an uncommon SNR value in gunshot detection systems.

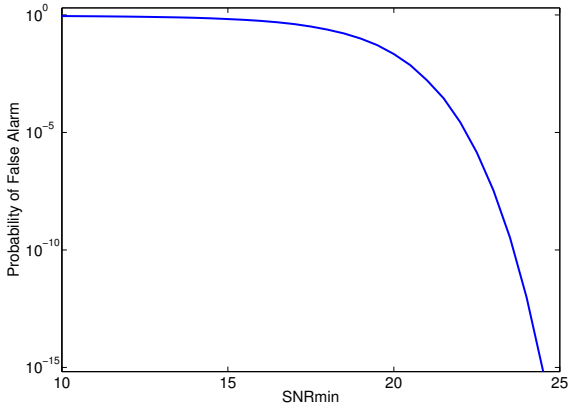

Fig. 1. Performance evaluation of event detection stage with $\text{SNR}_{\min}$.

4.3. Evaluating Gunshot Recognition Stage

We devised separate experiments to evaluate performance of cross-correlation maximum, LPC coefficients and MFCC features for gunshot recognition. Once a database of relevant feature sets corresponding to both

[1]The audio database we used to run our experiments is available at http://adcom.lums.edu.pk/gunshotdatabase.html

gunshot and outsider signals was built, eight-fold cross-validation [14] was used to train and test SVMs. In particular, a separate eight-fold cross-validation experiment was devised for LPC coefficients and MFCCs. Each dataset was divided into eight subsets. In turn, seven subsets were used to train the SVM classifier and the remaining one was used to test it. However, for fair comparison with [4], the cross-correlation feature was tested with a threshold only.

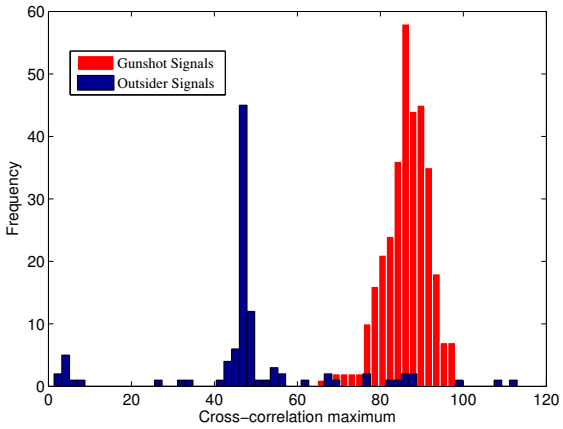

Fig. 2. Histogram of cross-correlation maxima of 434 audio clips with a gunshot template.

To compare the performance of the different feature sets, we used the average TPR and FPR as measures of classification accuracy. Average TPR is defined as the percentage of gunshots classified as gunshots in the testing subset as an average over the eight turns of the cross-validation test. Similarly, average FPR is the percentage of outsider signals classified as gunshots in the testing subset as an average over the eight turns of the cross-validation test. The results of the aforementioned three experiments are shown in Table 1. Though correlation based template matching has been proposed as a high performance gunshot detection algorithm in [7] [4], it gives a high FPR with our database. Moreover, Figure 2 shows the inseparable overlap in cross-correlation maxima of gunshots and outsider signals with a gunshot template. Thus, we need to use more feature(s) along with cross-correlation maximum to separate gunshots and outsider signals in the feature space. From Table 1, it can be seen that LPC coefficients give a relatively higher average TPR while correlation-based feature gives a relatively lower average FPR. The two features are then used together with SVMs to get a better average TPR and FPR simultaneously. The results indicate that even though we improve the TPR, the false alarm rate of 8.33% is still too high for practical purposes. So, we try out the kernel trick with our combinative feature set. In Table 2, we show that LPC coefficients, cross-correlation maximum and RBF kernel can be used together with SVMs to obtain a 97.6% TPR and a negligible average FPR [2]. For comparison, we display results for the other features too when used with the RBF kernel. For these experiments involving use of the RBF kernel, the kernel parameter and the penalty parameter were optimized using the grid-searching algorithm [15]; different pairs of kernel parameter and penalty parameter were tested and the one which gave the best detection performance in the eight-fold cross-validation was used to generate results. The use of cross-validation ensured that the classifier did not over-fit to the training data.

Feature Set	**Classifier**	**TPR**	**FPR**
Cross-correlation maximum [7]	Threshold	94.580	8.824
MFCC	SVM	97.321	50.000
LPC	SVM	99.702	11.458
LPC + Cross-correlation	SVM	99.702	8.333

Table 1. Classification accuracy (as percentages) for different feature sets with linear kernel and SVMs

Feature Set	**TPR**	**FPR**
MFCC	92.497	15.625
LPC	96.429	1.042
Cross-correlation maximum	97.917	10.417
LPC + Cross-correlation	97.619	~ 0

Table 2. Classification accuracy (as percentages) for different feature sets with RBF kernel and SVMs.

5. CONCLUSION

In this paper, we have proposed a gunshot detection system with high gunshot detection performance, robustness to noise and low computational complexity. Our proposed analytical framework for event detection enables minimization of computational complexity of the system without escalating the chances of missing out on acoustic events. Moreover, we showed that template matching cannot be used alone for gunshot classification because of the high FPR. Instead, we propose to use the template matching feature in conjunction with eighth order LPC coefficients and gaussian RBF kernel to train SVMs for gunshot detection.

[2] Of the 102 outsider signals we tested with LPC + Cross-correlation, we did not get any false positive.

6. REFERENCES

[1] R.C. Maher, "Acoustical characterization of gunshots," in *Signal Processing Applications for Public Security and Forensics, 2007. SAFE'07. IEEE Workshop on*. IET, 2007, pp. 1–5.

[2] L. Gerosa, G. Valenzise, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and gunshot detection in noisy environments," in *15th European Signal Processing Conference (EUSIPCO-07), Sep. 3-7, Poznan, Poland*, 2007.

[3] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis, "Gunshot detection in audio streams from movies by means of dynamic programming and bayesian networks," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*. IEEE, 2008, pp. 21–24.

[4] I.L. Freire and J.A. Apolinário Jr, "Gunshot detection in noisy environments," in *Proceeding of the 7th International Telecommunications Symposium, Manaus, Brazil*, 2010.

[5] A. Dufaux, *Detection and recognition of impulsive sounds signals*, Ph.D. thesis, Ph. D. thesis, Faculté des sciences de lUniversité de Neuchatel, 2001.

[6] L.G. Mazerolle, C. Watkins, D. Rogan, and J. Frank, *Random gunfire problems and gunshot detection systems*, US Department of Justice, Office of Justice Programs, National Institute of Justice, 1999.

[7] A. Chacón-Rodríguez, P. Julián, L. Castro, P. Alvarado, and N. Hernández, "Evaluation of gunshot detection algorithms," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 58, no. 2, pp. 363–373, 2011.

[8] X. Huang, A. Acero, H.W. Hon, et al., *Spoken language processing*, vol. 15, Prentice Hall PTR New Jersey, 2001.

[9] B. Schölkopf and A.J. Smola, *Learning with kernels: Support vector machines, regularization, optimization, and beyond*, MIT press, 2001.

[10] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523 – 531, april 1967.

[11] M. Abramowitz and I.A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55, Dover publications, 1965.

[12] JG Proakis, *Digital Communications*, McGraw-Hill, fourth edition, 2001.

[13] M. Slaney, "Auditory toolbox," *Interval Research Corporation, Tech. Rep*, vol. 10, pp. 1998, 1998.

[14] R. Kohavi et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *International joint Conference on artificial intelligence*. Lawrence Erlbaum Associates Ltd, 1995, vol. 14, pp. 1137–1145.

[15] C.W. Hsu, C.C. Chang, C.J. Lin, et al., "A practical guide to support vector classification," 2003.