GUN TYPE RECOGNITION FROM GUNSHOT AUDIO RECORDINGS

Eva Kiktova, Martin Lojka, Matus Pleva, Jozef Juhar, Anton Cizmar

Technical university of Kosice, Faculty of Electrical Engineering and Informatics
Department of Electronics and Multimedia Communications, Letna 9, 04120 Kosice, Slovakia
{eva.kiktova, martin.lojka, matus.pleva, jozef.juhar, anton.cizmar}@tuke.sk

ABSTRACT

This paper describes an extension of an intelligent acoustic event detection system, which is able to recognize sounds of dangerous events such as breaking glass or gunshot sounds in urban environment from commonly used noise monitoring stations. We propose to extend the system the way that it would not only detect the gunshots, but it would identify a suspects gun/pistol type as well. Such extension could help the investigation process and the suspect identification. The proposed extension provides a new functionality of the gun type recognition (classification) based on audio recordings captured. This research topic is discussed in other research papers marginally. Different kinds of features were extracted for this challenging task and feature vectors were reduced by using mutual information based feature selection algorithms. The proposed system uses two phase selection process, HMM (Hidden Markov Model) classification and Viterbi based decoding algorithm. The presented approach reached promising results in the experiments (higher than 80% of ACC and TPR).

***Index Terms*—** Gun type classification, acoustic event detection, feature selection, HMM

1. INTRODUCTION

Systems for detecting abnormal sound events are very popular in these days. Research in this area is demanding because of ethical and legislation issues of audio surveillance. During our work in the security related FP7 project INDECT [1] we introduced an idea of using noise monitoring stations being common in noisy urban environments [2], [3] (larger cities[1], or cities in airport/factory neighborhood[2]) for the sounds of dangerous abnormal events detection that could help the police forces to monitor possible threat activity in the city. The system could provide a gun type recognition for the forensic investigation of the crime scene if there was an audio recording available.

Audio signal captured by a far field noise microphone could be used for the audio monitoring because a low level human voice is being masked with a noise.

Such intelligent acoustic events detection system can be applied also in banks, airports, hospitals, jails, and in industrial or any other specific applications. It works autonomously and generates alert (and stores a log file) only when some abnormal situation is detected. Design of such system requires taking the specific aspects into account such as place of installation, appropriate models for events and backgrounds, connection to another equipment or system and, of course, the specific end-users requirements. In this paper, we use the term 'acoustic event' referring to the relatively short lasting, high level, randomly occurring and relevant sound.

An acoustic event recognition is a relatively new research topic in the comparison to a speech recognition. In [4] a detection and localization of acoustic events is described, in [5] an acoustic unit descriptors are used for an event detection and in [6] a hybrid learning approach is presented.

The acoustic event recognition can be also applied in smart rooms as it was described in [7] and also in the human assistant applications such as fall detection system being presented in [8]. There are some research papers related to guns, security or forensic applications, for example a user verification system based on a grip-pattern recognition for smart guns described in [9] or gun-type recognition based on the image reported in [10] but the recognition of gun/pistol type based on audio information is discussed insufficiently. Our paper fills a white space in this research field.

The detection system presented in this paper is compound of the two functional blocks, i.e. a front-end processing block and a classification block. The front-end processing plays a very important role in the whole system because its performance directly depends on the quality of extracted (possibly selected) features. All features in use should enhance the desired signal characteristics and also provide information that allows for a correct classification of input samples. A front-end processing includes these main operations:

- feature extraction,
- feature selection.

[1]On the implementation of the Environmental Noise Directive in accordance with Article 11 of Directive 2002/49/EC, where 163 cities with over 250 000 inhabitants were reported on http://noise.eionet.europa.eu/

[2]Airports with Noise Restrictions http://www.boeing.com/boeing/commercial/noise/list.page

978-1-4799-8105-2/15/$31.00

Different types of features can be extracted from an input sound [11], [12]. Spectral characteristics are used very often. Some of them produce scalar values (Spectral Flux, Spectral Roll-off, Audio Spectrum Centroid, Audio Spectrum Spread, etc.) and other types generate m-dimensional feature vector (Audio Spectrum Flatness - ASF, Audio Spectrum Envelope - ASE, log-filter bank coefficients - FBANK, Audio Spectrum Projection - ASP, etc.) A promising performance was achieved also with speech-inspired cepstral features like Mel-Frequency Cepstral coefficients - MFCC or Perceptual Linear Prediction coefficients - PLP. Temporal features are not usually used alone, but rather as a supplement to other features. The commonly used temporal features are Zero Crossing Rate - ZCR, Audio Waveform - AW (down-sampled envelope), energy, statistical parameters (Skewness, Kurtosis), etc.

After successful feature extraction an optional feature selection process can follow. Two main selection approaches are reported in the literature [13]: wrapper using classifiers [14], [15] and filter selection, that is independent from the classifier [16]. Identification of an appropriate selection criterion is the crucial aspect for effective filter based selection. An improvement of the system performance is then reached by elimination of redundant features (also called "noise features"). The filter selection approach based on the mutual information was applied in this work. The used algorithms are described in the Sec. 2.2.

Feature transformation methods [17] such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) can be also performed for dimension reduction of the feature vectors [18].

Our work is based on the Hidden Markov Models - HMM [12], [19] that represent a compact characteristic of analyzed sounds. The system solution proposed by us is focused on the security or audio surveillance (urban environment, banks, airport, etc.) applications. Its first phase is dedicated to abnormal sounds (gun shots and breaking glass) detection. After the successful gunshot detection the gun type classification phase follows. It is possible to recognize four types of gun in this phase.

The first subtask "a correct detection of abnormal sounds" was precisely described and evaluated in our previous work [20]. We adopted and extended this approach for the second subtask i.e. the gun type recognition. This issue is presented in this paper.

2. FEATURE EXTRACTION AND SELECTION

A method of the recognition improvement is combining multiple features in a vector and reducing the vector dimension by performing an effective selection that eliminates the redundant or irrelevant data. This idea was applied for the gun type recognition.

2.1. Feature extraction

In this work ZCR, Flux, Roll-off, ASF, ASE, ASS, ASC, FBANK features with energy, delta and acceleration coefficients are used and two candidate feature sets (CS1 and CS2) are created [20]:

- CS1: ZCR + Flux + Roll-off + ASF + ASS + ASC + FBANK_EDA (dim 61),
- CS2: ZCR + Flux + Roll-off + ASE + ASS + ASC + FBANK_EDA (dim 82).

These candidate feature sets are used for the selection process inspired by the mutual information based algorithm. More information about feature extraction can be found in [20].

2.2. Feature selection theory

Mutual information [21], [22] is a measure of dependence between random variables.

Two random variables x and y and their *mutual information* $I(x;y)$ is defined as follows:

$$I(x,y) = \sum_{i,j} p(x_i,y_j) log_2 \frac{p(x_i,y_j)}{p(x_i)p(y_j)}, \tag{1}$$

where the mutual information I of two variables x and y is based on their joint probabilistic distribution $p(x,y)$ and their probabilities $p(x)$ and $p(y)$. The feature selection according to the *mutual information* is implemented in the MRMR MIBASE algorithm.

The concept of mutual information [21], [22] can be easily expanded to include more than two random variables. According to the chain rule the *joint mutual information* $I(x_1,...,x_n;y)$ between a set of features $(x_1,...x_n)$ and y is defined as follows:

$$I(x_1,...,x_n;y) = \sum_{i=1}^{n} I(x_i;y|x_{i-1},x_{i-2},...x_1). \tag{2}$$

JMI (Joint Mutual Information) describes the amount of decrease in an uncertainty of variable y achieved by using an information provided by the feature vector $(x_1,...x_n)$.

Min-Redundancy Max-Relevance (MRMR) algorithms [23], [24] select a compact set of superior features by applying a combination of two independent criteria: *maximal relevance* and *minimal redundancy.*

In the of the first criterion *maximal relevance* features are selected according to the highest relevance to (dependency on) the target class c. The relevance can be interpreted as correlation or mutual information (Eq.1) defining the dependencies between variables.

Maximal relevance (maximal dependency - $maxD$) of the feature set S containing features x_i can be described by formula:

$$maxD(S,c), \qquad \text{where} \qquad D = \frac{1}{|S|^2} \sum_{x_i \in S} I(x_i,c), \tag{3}$$

where $maxD(S, c)$ is computed using mean of all mutual information values between the individual feature x_i and the corresponding c class.

Features selected by *maximal relevance* usually have a rich redundancy (strong dependency between features). When two features x_i, x_j are heavily dependent on each other, their class-discriminant power would not change much if one of them was removed. For this reason the *minimal redundancy* $(minR)$ criterion is applied by formula:

$$minR(S), \qquad \text{where} \qquad R = \frac{1}{|S|^2} \sum_{x_i x_j \in S} I(x_i, x_j). \quad (4)$$

MRMR combines these two criteria *maximum relevance* and *minimum redundancy* into the single criterion function, which shall optimize D and R simultaneously. Regarding this, two simple functions are considered:

$$MID = max(D - R), \quad (5)$$

$$MIQ = max(D/R). \quad (6)$$

The Eq. 5 describes MRMR feature selection scheme based on the Mutual Information Difference (MID) and Eq. 6 describes the Mutual Information Quotient (MIQ) selection scheme.

3. EXPERIMENTAL SETUP

In this section the information about the performed selections and the sound database with respect to the experimental work is presented.

3.1. Feature selection

Several selections based on the mutual information such as (MRMR_MIBASE/MID/MIQ) and the joint mutual information (JMI) selection criteria were performed:

- CS1 (dim 61) reduced to dim 30, 40 by MIBASE, MIQ, MID and JMI,
- CS2 (dim 82) reduced to dim 30, 40 by MIBASE, MIQ, MID and JMI.

The selection was also applied on the baseline feature sets MIBASE_CS1_30 and MIQ_CS2_40, which achieved very promising results in this new task (gun type recognition). Two feature sets mentioned above were also effective for the baseline AED system, where they achieved the detection score ACC and F-measure over 99% [20].

In this phase there were three best selection criteria MIBASE, MIQ and JMI used:

- MIBASE_CS1_30 reduced by MIBASE, MIQ and JMI (dim 22, 24, 26, 28),

Gun	cartridge	train	test
Pi vz.82	9x18mm / 9mm Makarov	137	19
1911	.45 Auto / .45 ACP	78	9
Ruger LCP	.380 Auto / .380 ACP / / 9x17mm / / 9mm Browning Short	78	7
Alarm pistol	-	79	8
Total		372	43

Table 1. Used part of the sound database.

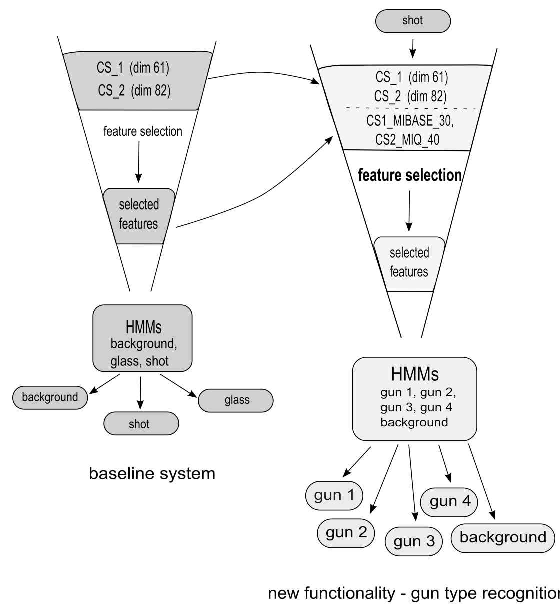

Fig. 2. The principal scheme of AED system supporting gun type recognition.

- MIQ_CS2_40 reduced by MIBASE, MIQ and JMI (dim 22, 24, 26, 28, 30, 32, 34, 36, 38).

This approach uses a two-phase selection. In the first phase features are selected for the acoustic event/ background and in the second phase features are reselected for the particular gun type, i.e. gun1/gun2/gun3/gun4.

3.2. Sound database

The sound database used in this work (see Tab.1) is a part of JDAE TUKE [25]. It was mainly designed for the acoustic event detection task, but it can also be used for the gun type recognition. It contains various types of weapons, but only four types of them have more than 80 instances. This is the reason why we recognize only four types of them Fig. 1.

Fig. 1. Used guns (from left Pi vz.82, 1911, Ruger LCP and alarm pistol).

4. SYSTEM DESCRIPTION

The acoustic event detection system described in works [26], [27], [28], recognizes dangerous sounds and generates alert if gunshot or braking glass sound is detected. If a gunshot sound is detected, a new gun type recognition system will try to identify the particular gun type. The principal system scheme is depicted in Fig.2.

For HMM training and testing, the HTK environment was used [12]. A decoding process was based on the Viterbi algorithm. For the evaluation of the overall recognition (background and four types of guns) the Accuracy (Acc%) [12] was used:

$$ACC\,[\%] = \frac{N - D - S - I}{N} \times 100, \tag{7}$$

where D is the number of deletion errors, S is the number of substitution errors, I is the number of insertion errors, and N is the total number of labels in the reference [12].

For the gun type recognition a True Positive Rate (TPR%) [29] is computed. This measure reflects only hit rate for the considered guns:

$$TPR\,[\%] = (TP/N) \times 100, \tag{8}$$

where TP is the number of correct gun recognitions and N is the total number of labels in the reference. In this TPR evaluation, the background is not taken into account.

We created and evaluated each feature set adopted from [20] and all new selected sets (as was reported in 3.1) by different types of HMM (from one to four states and from 1 to 1024 Probability Density Functions - PDF). Models were created for the ambient sound and four types of guns (Pi vz.82, 1911, Ruger LCP and alarm pistol). Evaluations of individual feature kinds were also performed, but the recognition results were not satisfactory because only 80.61% Acc and 76.74% TPR were reached in the case of FBANK_EDA 3/256. The rest of features (ASE, ASF, ASC, ASS, Flux, Roll-off, ZCR) achieved worse recognition results.

We decided to find feature sets that would be able to recognize a gun type more accurately than 80% for both evaluation measures. An exhaustive search was performed and finally we identified feature sets that were capable of that.

5. RESULTS

The best results achieved for the gun type recognition are depicted in the Fig. 3 (more than 80% for both measures). The first group of results belongs to the two-phase selection approach, the second group presents results associated to the directly selected sets and the last group is associated to the results of the baseline system sets. The most results depicted in the Fig. 3 belong to the two-phase selection strategy (i.e. combination of two selection criteria). Generally CS2 achieved better results than CS1. In the most cases three-state models with 128 and 256 PDFs brought desired results. The core of successful feature sets was based on the CS2 including mainly static and first temporal derivations of FBANK coefficients, ASC, ASS and Flux. On the other hand, Roll-off and ZCR features were not used at all. Less accurate classifications were caused by ASF features, which were in candidate sets derived from CS1. According to the applied selection algorithms the feature FBANK_4 was proposed as the most important feature for both candidate sets.

6. DISCUSSION

Our experiments were limited with relatively small amount of train and test data available. We would like to extend our database, split it into the balanced training, development and test data sets, while each tested gun type would have the same number of shot sound instances if there will be a possibility of cooperation with gun/pistol owners/sellers. This way our results will be more objective. Recordings used in the experiments were produced by one physical gun/pistol per type, therefore we suggest to perform also experiments with recordings of several guns of the same type.

7. CONCLUSION

In this work, the gun/pistol type recognition taking place after gunshot sound detection was described. The main attention was paid to the feature selection by MRMR (MIBASE, MID, MIQ) and JMI that ensured the improvement of recognition results by the elimination of irrelevant features. Selected features highlighted the nature of a sound analyzed. We also

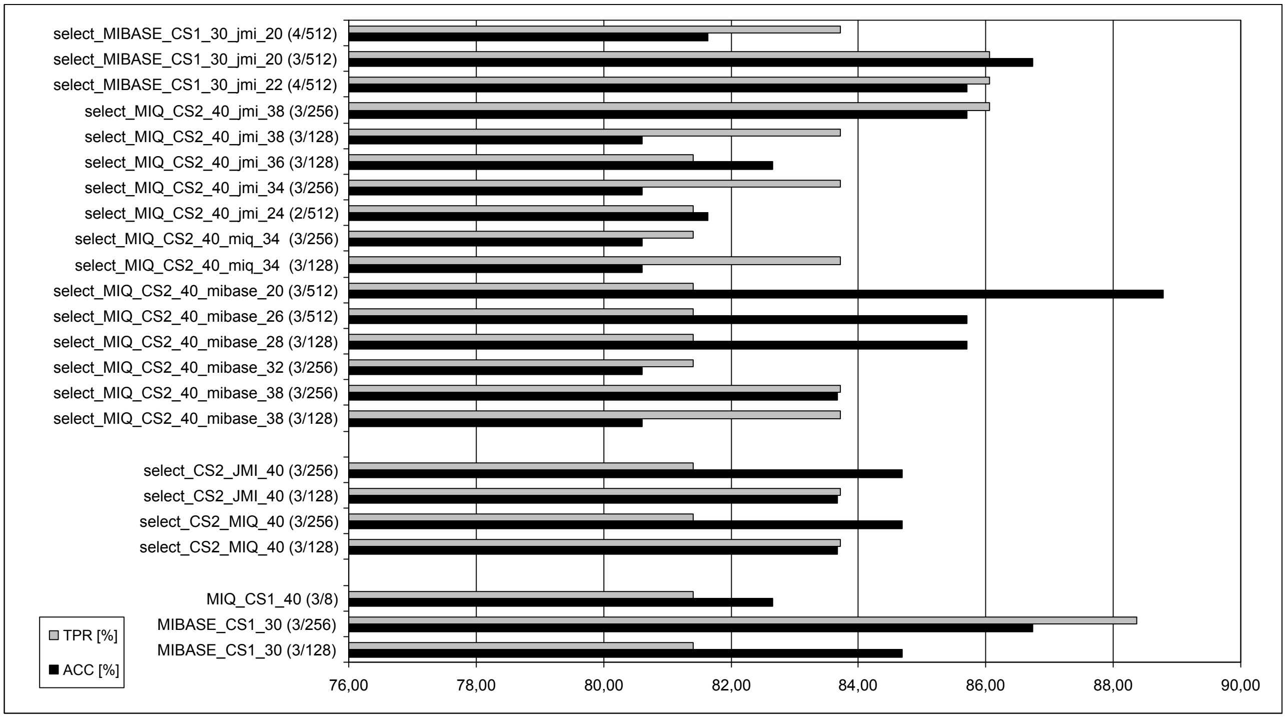

Fig. 3. Results of the gun type recognition. Denotation: feature set acronym, its dimension, type of HMM (state/PDF).

explored the relation between the optimal feature subset selection and the relevance of selected features. The process of two-phase selection was successfully applied and brought promising results for such systems.

Our system is currently limited to identify four gun types, but the proposed methodology and the approach that follows the system could be extended for detecting wide range of types of guns currently in use and also to other classes of specific sounds (walk, stair walk, etc.) reliable for identification of the suspect.

Through the use of selection algorithms we identified several combinations (feature sets and particular HMMs) that have produced recognition results of ACC and TPR higher than 80%.

Our system introduces also a very useful tool for the forensic investigation because the exact time of the gunshot and a detected gun type could be sent on-line to the police forces, possibly helping in the investigation process.

8. ACKNOWLEDGMENTS

This paper is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 26220220182, supported by the Research & Development Operational Programme funded by the ERDF.

9. REFERENCES

[1] "http://www.indect-project.eu/," FP7 project website. Last accessed Jun, 2014.

[2] L. Hong and etc., "Sound source study in Shenzhen China," in *Proc. INTER-NOISE*, 2014, pp. 1–4.

[3] F. Van der Eerden, F. Graafland, P. Wessels, A. Segers, and E. Salomons, "Model based monitoring of traffic noise in an urban district," in *Proc. INTER-NOISE*, 2014, pp. 16–19.

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

[5] A. Kumar, P. Dighe, R. Singh, S. Chaudhuri, and B. Raj, "Audio event detection from acoustic unit occurrence patterns," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2012, pp. 489–492.

[6] R. Radhakrishnan, A. Divakaran, and P. Smaragdis, "Audio analysis for surveillance applications," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2005, pp. 158–161.

[7] T. Butko, F G. Pla, C. Segura, C. Nadeu, and J. Hernando, "Two-source acoustic event detection and localization: Online implementation in a smart-room," in *Proceedings of the 17th European Signal Processing Conference (EUSIPCO11), Barcelona, Spain*, 2011, vol. 29, pp. 1317–1321.

[8] X. Zhuang, J. Huang, G. Potamianos, and M. Hasegawa-Johnson, "Acoustic fall detection using Gaussian mixture models and GMM supervectors," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, 2009, pp. 69–72.

[9] J. A. Kauffman, A. M. Bazen, S. H. Gerez, and R. N. J. Veldhuis, "Grip-pattern recognition for smart guns," in *Proceedings of ProRISC 2003, 14th Annual Workshop on Circuits, Systems and Signal Processing*, 2003, pp. 379–384.

[10] K. M. L. Kasemsan, "The classification of guns type using image recognition theory," *International Journal of Information and Electronics Engineering*, vol. 4 (1), pp. 54–58, 2014.

[11] K. Hyoung-Gook, N. Moreau, and T. Sikora, *MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval*, Wiley & Sons, 2005.

[12] S. Young and etc., *The HTK Book, version 3.4*, Cambridge University Engineering Department, Cambridge, UK, 2006.

[13] I. Kojadinovic and T. Wottka, "Comparison between a filter and a wrapper approach to variable," in *ESIT 2000, 14-15 September 2000, Aachen, Germany*, 2000, pp. 311–321.

[14] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *ARTIFICIAL INTELLIGENCE*, vol. 97, no. 1, pp. 273–324, 1997.

[15] D. Hladek, J. Vascak, and P. Sincak, *Towards Fuzzy Learning Classifier Systems: Theory and Application of the Reinforcement Learning, Fuzzy Logic and Learning Classifier Systems*, LAP LAMBERT Academic Publishing, 2012.

[16] N. Sanchez-Marono, A. Alonso-Betanzos, and M. Tombilla-Sanroman, "Filter methods for feature selection - a comparative study," in *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, vol. 4881 of *Lecture Notes in Computer Science*, pp. 178–187. Springer Berlin Heidelberg, 2007.

[17] J. Juhar and P. Viszlay, "Linear feature transformations in Slovak phoneme-based continuous speech recognition," in *Modern Speech Recognition Approaches with Case Studies*. 2012, pp. 131–154, InTech Open Access.

[18] E. Vozarikova, P. Viszlay, J. Juhar, and A. Cizmar, "Acoustic events detection via PCA-based feature extraction," *Journal of Computer Science and Control Systems*, vol. 3 (2), pp. 255–258, 2010.

[19] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 4, no. 2, pp. R13, 2007.

[20] E. Kiktova-Vozarikova, J. Juhar, and A. Cizmar, "Feature selection for acoustic events detection," *Multimedia Tools and Applications*, pp. 1–21, 2013, Article in Press.

[21] H. Hua Yang and J. John Moody, "Feature selection based on joint mutual information," in *In Proceedings of International ICSC Symposium on Advances in Intelligent Data Analysis*, 1999, pp. 22–25.

[22] G. D. Tourassi, E. D. Frederick, M. K. Markey, and C. E. Floyd, "Application of the mutual information criterion for feature selection in computer-aided diagnosis," *Medical Physics*, vol. 28, no. 12, pp. 2394–2402, 2001.

[23] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[24] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," in *Bioinformatics Conference, 2003. CSB 2003. Proceedings of the 2003 IEEE*, 2003, pp. 523–528.

[25] M. Pleva, E. Vozarikova, L. Dobos, and A. Cizmar, "The joint database of audio events and backgrounds for monitoring of urban areas," *Journal of Electrical and Electronics Engineering*, vol. 4 (1), pp. 185–188, 2011.

[26] M. Lojka, M. Pleva, E. Kiktova, J. Juhar, and A. Cizmar, "EAR-TUKE: The acoustic event detection system," in *Int. conf. Multimedia Communications, Services and Security, MCSS 2014: Communications in Computer and Information Science*, 2014, vol. CCIS 429, pp. 137–148.

[27] E. Vozarikova, M. Pleva, J. Juhar, and A. Cizmar, "Surveillance system based on the acoustic events detection," *Journal of Electrical and Electronics Engineering*, vol. 4 (1), pp. 255–258, 2011.

[28] M. Pleva, M. Lojka, and J. Juhar, "Modified Viterbi decoder for long-term audio events monitoring," *Journal of Electrical and Electronics Engineering*, vol. 5 (1), pp. 195–198, 2012.

[29] N. A. Macmillan and D. C. Creelman, *Detection Theory: A User's Guide*, Lawrence Erlbaum, 2 edition, 2004.