Pattern Recognition Neural Network as a Tool for Pest Birds Detection

Petr Dolezel
Faculty of Electrical Engineering and Informatics
University of Pardubice
Pardubice, Czech Republic
Email: petr.dolezel@upce.cz

Pavel Skrabanek
Faculty of Electrical Engineering and Informatics
University of Pardubice
Pardubice, Czech Republic
Email: pavel.skrabanek@upce.cz

Lumir Gago
Faculty of Electrical Engineering and Informatics
University of Pardubice
Pardubice, Czech Republic
Email: lumir.gago@student.upce.cz

***Abstract*—Various kinds of vermin have been considered as a huge problem since primeval times. Over this period, means of protection against vermin have developed to be very quick and efficient. However, new goals in protection have appeared recently which reflects legislative changes in most countries. Public opinion has shifted towards greater environment protection. Nowadays, vermin control systems have turned from being used globally into local applications and from being applied preventively into casual usage. Thus, accurate vermin detection units are becoming very important parts of vermin control systems. This situation is valid in agricultural areas (e.g. vineyards) which are protected against pest birds, too. Reflecting on the current situation, a feedforward multilayer artificial neural network, aimed on detection of European starling in vineyards, is presented in this paper. Except a description and validation of the detection method, the idea of the comprehensive protection system is also outlined in this paper.**

I. Introduction

A need to protect vineyards against various pests was obvious since the very beginning of viticulture. Considering the amount of damages caused by birds, they are perceived to be particularly undesirable vermins. Their legal stratus adds further complications for the viticulture. Therefore, various technologies aimed on protection of crop, such as mist nets, propane cannons, or sound alarms, have been invented in the last decades. An extensive summary of these techniques can be found in [1]. Although there are a number of techniques to choose from, the most effective ones are either too expensive to be efficiently applied, or illegal these days. Let us illustrate the current situation on the example of well-known viticultural region of South Moravia, Czech Republic. In 2014, only two means of protection were applied: the propane cannons and gas guns. The noise pollution produced by both techniques (115 – 130 decibels) is really annoying for residents living near vineyards. Indeed, propane cannons are operated continuously from the second half of August till the beginning of November.

On the initiative of the authors of this paper, in cooperation with several vineyards situated in South Moravia, a more sophisticated system for pest bird control is being developed. The hazing system is designed as a building kit which allows its appropriate installation just according to circumstances. The kit consists of a Central Control Unit, a Detection Unit and a Scaring Unit. A schematic representation of a possible placement of the units is shown in Fig. 1 while the data flow schedule is presented in Fig. 2.

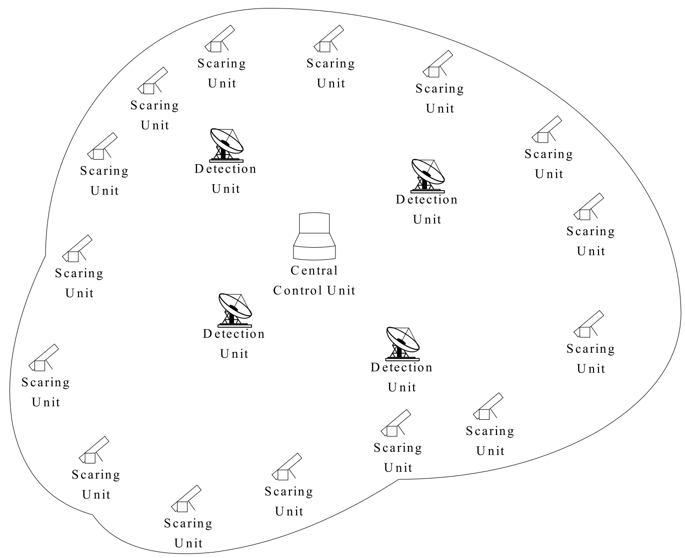

Fig. 1. A layout of a possible installation of the system

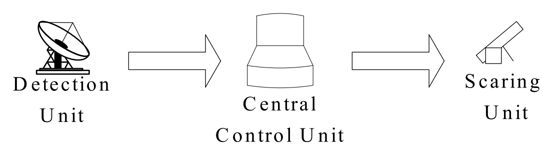

Fig. 2. Data flow between the blocks

Clearly, the whole system of protection is activated just after the flock of pest birds is detected in a particular area of a vineyard. Thus, from this point of view, the crucial part of the system is the detection unit. So far, several techniques, aimed on detection and recognition of birds, have been introduced. They use either audio data [2], images [4], [5], [6], or radio waves [3] as the input signal. Combination of these signals is also possible [7]. Many of these solutions are applicable on a wide range of bird species; however, their universality is also their weakness. Indeed, achievement of high detection reliability for an extensive range of the species is really not a simple task. From this perspective, solutions aimed on a few species are more appropriate. In the case of the Moravian vineyards, the majority of damages are caused by the European starling (Sturnus vulgaris). The prior knowledge of the target species has allowed us to develop an appropriate solution. This solution carries out: a) diagnose of a presence of starlings in a vineyard, b) localization of a position of a flock as exactly as possible.

II. Problem formulation

As is shown in Fig. 1, detection units are spread over an protected area. A comparison of the outputs provided by detection devices allows estimation of the position of a flock. Such an approach is described e.g. in [8]. As was stated in the introduction part, three types of input signal can be considered for recognition and localization of birds. In our solution, field sound samples are used as a source of information. This selection has been taken after a solid consideration including extensive literature research [9], [10], [11], [12]. Indeed, use of the audio data brings following advantages: a) it is less computationally demanding than image processing or signal processing of radio waves; b) no visual contact between an emitter (a bird) and a receiver (the unit) is required. However, birds produce a wide variety of sounds, e.g. songs, calls, or mechanical sounds. Naturally, every type of the sound should be taken into account by the pest bird detection.

A data flow in the proposed detection unit is illustrated in Fig. 3. The solution for each block in the pipeline is described in the following sections.

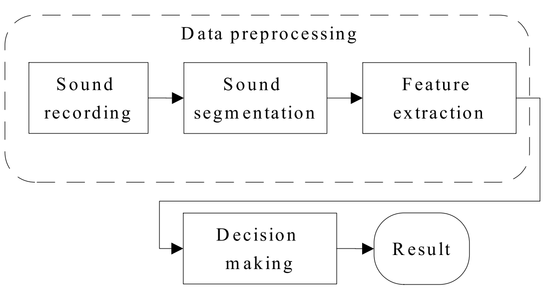

Fig. 3. Flow chart of the detection device

III. Data preprocessing

In our conception, the term data preprocessing covers: data acquisition, sound segmentation and feature extraction, where these operations are performed in this order. Thus, the output of the data preprocessing is a feature vector.

A. Sound recording

A field device used for reception of the input signal considerably influence a resulting performance of the overall solution. Within this paper, single-channel (mono) sounds with sample rate of 44100 Hz and double precision are used. A typical signal gained by the field device is shown in Fig. 4.

B. Sound segmentation

The input signal has to be properly modified before the feature extraction is performed. In our solution, the signal is divided into segments of constant length. Problematic part of the design of this step is selection of the segment length. As follows from our survey [13], [14], [15], there is no approach to determine this value analytically. The values of the length may vary from 10 ms till 10 s. Clearly, the pertinent segment length depends on the selected feature extraction technique. In this particular case, the length of the segment is determined experimentally which will be described in the next sections.

C. Features extractions

At the feature extraction level, the most significant attributes are separate from the segmented signal. Generally, the feature extraction is often considered to have the greatest impact on the resulting performance of an overall solution. For the audio signal, many feature extraction algorithms have been already published. Let us mention at least Fast Fourier Transform [16], Linear Prediction Coding [17], Perceptual Linear Prediction [18], or Mel Frequency Cepstral Coefficients [19].

As a first engineering approach, Linear Prediction Coding (LPC) has been chosen. The LPC is suitable for modeling the vocalizations of birds by source-filter interaction [20]. It has also successfully proven in sound recognition, identification, and compression. Moreover, this method is not computationally demanding. Thus, the LPC is appropriate for the field devices where only simple microprocessors are used.

The basic idea of the LPC is that a current sound sample $s(n)$ can be closely approximated as a linear combination of past samples according to

$$\overline{s}(n) = \sum_{k=1}^{p} \alpha_k s(n-k), \tag{1}$$

where α_k is the k-th LPC coefficient, p is the number of the LPC coefficients, and $\overline{s}(n)$ is the approximation of the n-th sound sample $s(n)$.

To determine the values of the LPC coefficients, a prediction error

$$e(n) = s(n) - \overline{s}(n), \tag{2}$$

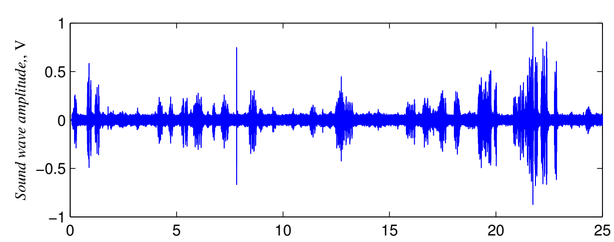

Fig. 4. Signal gained from the device (1102500 values for 25 seconds)

has to be minimized over the whole sound segment, where the cost function is given as

$$E = \sum_{n=1}^{N} e^2(n). \tag{3}$$

The variable N is the number of the samples in the current sound segment.

The minimum value of E occurs when the derivative is equal to zero [21] with respect to each of the parameters α. Such a way, a set of p equations is obtained:

$$\begin{array}{ccccccc} r(0)\alpha_1 & + & \dots & + & r(p-1)\alpha_p & = & -r(1), \\ r(1)\alpha_1 & + & \dots & + & r(p-2)\alpha_p & = & -r(2), \\ \vdots & + & \ddots & + & \vdots & = & \vdots \\ r(p-1)\alpha_1 & + & \dots & + & r(0)\alpha_p & = & -r(p), \end{array} \tag{4}$$

where $r(i)$ is the autocorrelation value computed as

$$r(i) = \sum_{m=0}^{N-1-i} s(m)s(m+i). \tag{5}$$

The advantage of this approach is that such a system of linear equations can be efficiently solved using the Levinson-Durbin algorithm [22]. The computational complexity of this algorithm is very low.

The prediction error $e(n)$ decreases inherently with increasing model order (number of α_k). On the other hand, a model as simple as possible, is demanded for computational complexity reduction. Anyway, the limited number of LPC coefficients closely represents even thousands of samples. Although there are some empiric recommendations, how to choose the suitable model order [20], it is determined here on the basis of the experimental results - see next sections.

IV. Decision making

For the decision making, several approaches are suggested. A nice example of the fuzzy sets theory used for decision making is described in [23]. On the other hand, support vector machine classifiers are successfully applied in [24] and expert systems are used in [25]. However, we decided to use the neural networks methodology, since neural networks are able to solve even linearly nonseparable and non-convex problems (in comparison to support vector machines) and their design is more generic compared to fuzzy and expert systems.

The neural network for decision making should be able to recognize some patterns in input data and to identify and categorize them. In our solution, juts two categories are considered: positive (target pest birds are present) and negative (they are not present). According to [26], hyperbolic tangent activation functions are recommended to be used in hidden layers while softmax activation functions are considered to be appropriate for the output layer. This topology of the feedforward network is called the pattern recognition network (PRN). The PRN can be generally represented as is shown in Fig. 5.

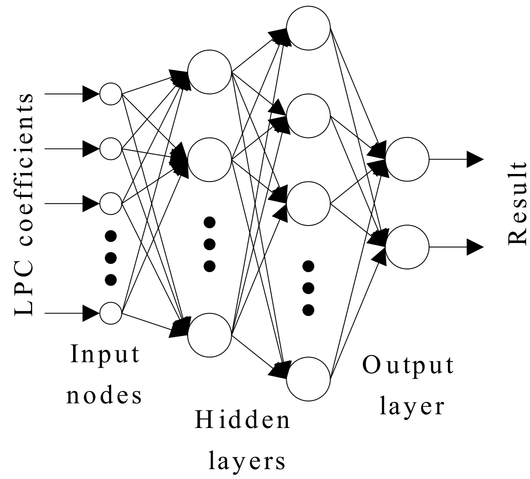

Fig. 5. General diagram of used neural network

The design of the PRN involves various activities. At first, appropriate input data has to be acquired. Afterwards, training, validation, and pruning can be performed. All these activities are described to the extent necessary in following subsections. More details about this process can be found e.g. in [27], [28].

A. Data acquisition

The training and testing sets consist of songs, calls and other sounds produced by more than 30 bird species. The total length of the sound records is about 50 minutes, where 6 minutes were produced by the European starling. Most of these records are general calls and alarm calls, which are the most common sounds produced by starlings in vineyards while feeding. Before the sets have been formed, the original data were modified. Specifically, silent segments were removed, the energies of all the recordings were balanced (see Fig. 6), and the noise was removed. Finally, the processed recordings were divided into samples of the defined length.

B. Training data

A source set of data consist of labeled feature vectors where the LPC with the appropriate setting was used by its forming. Just as reminder, two classes of samples are expected, where samples with starling sounds belong to the class positive, and the sounds of other species fall to the class negative. Within the training-validation process, 70% of the samples belong to the training set, 15% of them is placed into the testing set, and remaining 15% is used for the validation. The samples are always split up into these sets randomly.

C. Segments length and number of LPC coefficients

As mentioned in previous sections, the proper sound segment length, as well as the suitable number of LPC coefficients for features extractions, is needed to be determined. Since there are only empirical recommendations in related literature, we decided to perform a set of experiments to determine them. To be specific, the redundant PRN is trained periodically using the scaled conjugate gradient backpropagation [29], with

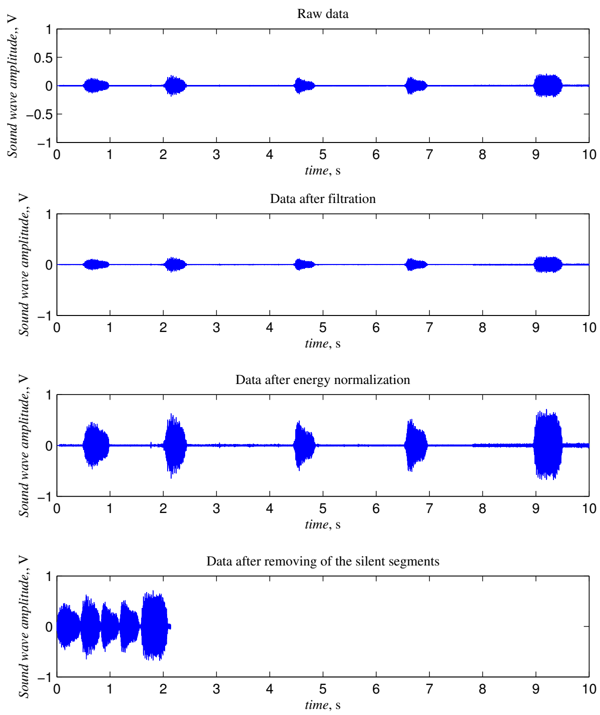

Fig. 6. Data processing

data gained from sounds divided into the segments of various lengths, with features extracted using various number of LPC coefficients. For each length, spread on the interval [10 ms; 6 s], and for each number of LPC coefficients from 5 to 20, 3000 training procedures with initial weights defined according to the Nguyen-Widrow initialization method [30] are performed. Statistical data of the validation process is observed. A validation error is defined using a cross entropy function, since it heavily penalizes outputs that are extremely inaccurate, with very little penalty for fairly correct classifications. For decision making, this error function is far more suitable than a classical mean square error.

$$E_{val} = -\frac{1}{N}\sum_{i=1}^{N}[o(i)\ln(y(i)) + (1-o(i))\ln(1-y(i))] \quad (6)$$

where $o(i)$ is the desired output, $y(i)$ is the actual output of the neural network and N is the amount of data. Data used for validation remained separate from the training and testing set.

Minima of the error function (6) for each combination of segment lengths and numbers of LPC coefficients (best of 3000 attempts) are shown in Fig. 7.

Looking at Fig. 7, a wide range of intervals is acceptable, since the error function value $\approx 10^{-8}$ is more than suitable as the performance index. Thus, considering also the computational complexity, the segment length is set to 3 s and the number of coefficients is set to 10.

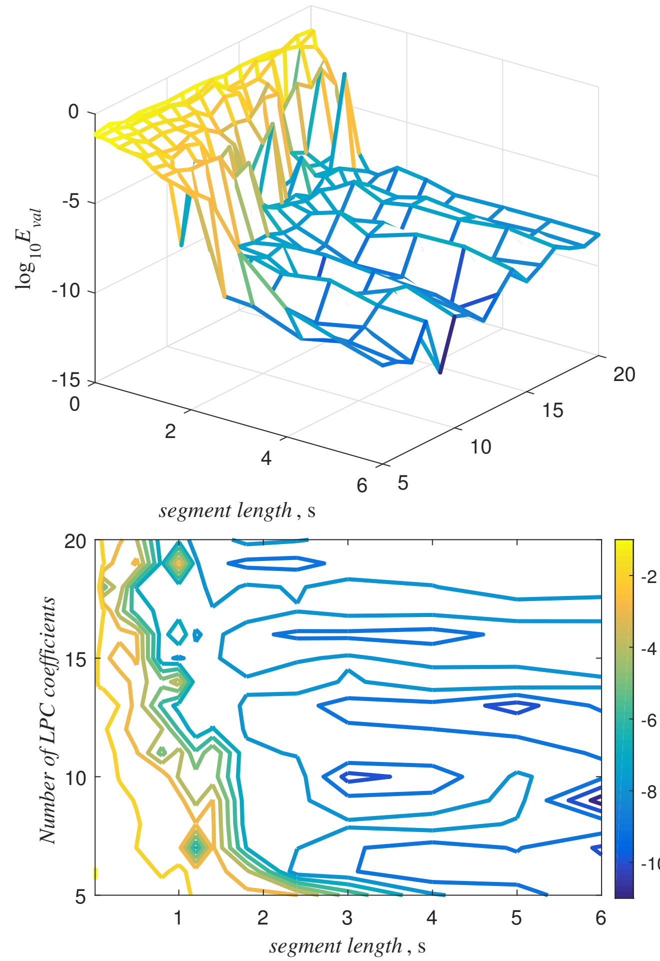

Fig. 7. The surface of the final value of the error function

D. Neural network training and pruning

As is well known, the training process seeks to find optimal setting of weights and biases. The aim of the pruning is to convert the net into a simpler one while keeping the performance of the original network. Both techniques are considered in order to determine an optimal topology of the PRN. For that purpose, PRNs of various topologies are trained 3000× using a scaled conjugate gradient algorithm. The training, testing and validation sets are randomly formed according to the specifications from subsection IV-B. The initial setting of the weights is given by the used Nguyen-Widrow initialization method. The obtained results are displayed using box graphs in Fig. 8. The central marks of the box graphs are medians, the edges are then 25^{th} and 75^{th} percentiles. The whiskers show the most extreme data points.

Looking at Fig. 8, a minimal value of the error function is achieved using the topology with two hidden layers, each with three neurons. In addition, this topology also provides a low median and variance of the resulting values. Thus, it is used

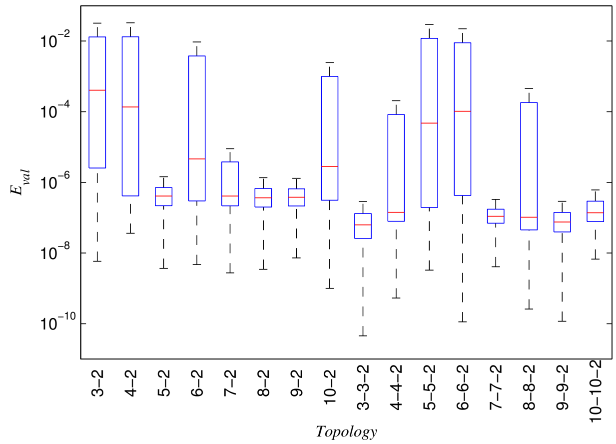

Fig. 8. Neural network pruning (X-axis represents the network topology to be examined)

for further work.

V. Validation of the Resulting Neural Network

According to the data in the previous section, 3-3-2 topology (10 inputs, 3 neurons in the first hidden layer, 3 neurons in the second hidden layer and 2 outputs) is declared as suitable. Such a network is validated again with new data divided into several groups according to the type of the input sound. Clearly, the ideal detection unit should recognize all types of starling sounds.

Mostly, it is common practice to use accuracy as the primary performance criterion. However, this single measure may not be sufficient enough [31]. Thus, two additional metrics, precision and recall, are proposed to evaluate the detection unit. The metrics are described by following equations:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{7}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{9}$$

where TP (true positive) is the number of correctly classified positive sounds, FN (false negative) is the number of misclassified positive sounds, FP (false positive) is the number of misclassified negative sounds, and TN (true negative) is the number of correctly classified negative sounds.

As validation data, starling general calls (5 minutes), alarm calls (4 minutes), begging calls produced by juvenile individuals (5 minutes) and songs (5 minutes) are used as positive samples; sounds produced by various other birds are used as negative ones. For each set of evaluation data, the same number of positive and negative samples is applied. In Table I, the results are shown.

TABLE I
Validation results - various kinds of birds

	Calls	Alarm calls	Begging calls	Songs	Total
Accuracy	0.9043	0.9516	0.8857	0.8627	0.8963
Precision	0.9362	0.9375	0.9355	0.9400	0.9375
Recall	0.8800	0.9677	0.8286	0.8103	0.8621

TABLE II
Validation results - natural sounds at vineyards

	Calls	Alarm calls	Begging calls	Songs	Total
Accuracy	0.9400	0.9841	0.9155	0.9000	0.9302
Precision	1.000	1.000	1.000	1.000	1.000
Recall	0.8800	0.9677	0.8286	0.8103	0.8621

As a second validation experiment, negative samples from the first experiment are replaced by natural sounds recorded in various vineyards (clearly, with no presence of starlings). Indeed, this experiment is less demanding on the detection unit, though it is closer to the field conditions of the actual usage. For this case, the results are shown in Table II.

The rates in view seem to be very promising (almost 90 % of correct answers in the first experiment, more than 93 % in the second one). In comparison to some works which deal with similar problematics, the rate is above average. In [32], the results are similar or worse, according to the used technique. In [33], even though the experiments cannot be compared directly to our approach, in general, the rates published there are lower than ours. Dealing with the information published in [34], note that the most important quantity, considering the usage of the detection unit, is the precision. And in our second experiment, precision equal to 1.000 is achieved. Thus, no false positive answer is gained. Clearly, there are some false positive answers gained in the first experiment. However, the data here is extracted from the birds sounds. Thus, the activation of the system may not be far from eligible. Therefore, this paper provides more reliable and robust solution than the approach published in [34].

VI. Conclusion

In this paper, a pattern recognition network, aimed on detection of European starling, was presented. The network was trained on the labeled set of field sounds. For the feature extraction, linear prediction coding was chosen for the following reasons: a) the LPC is suitable for modeling the vocalizations of birds, b) it has already proven by sound recognition, c) it is not computationally demanding. The last fact is especially important since the field devices are equipped only by simple microprocessors.

The proposed solution was subjected to tests. The obtained results indicate that the developed neural network is capable of detecting the European starling. The validation data shows very low level of false positive answers. Indeed, precisions for all the types of sounds are very high. This metric is the most important quality of the detection (false positive answers bring unnecessary noise pollution). Note that the qualities of the detection unit for begging calls and songs are similar to others. This interesting feature is achieved; though these types of sounds are very rarely present in the training set.

ACKNOWLEDGMENT

The work has been supported by the funds of the IGA, University of Pardubice, Czech Republic. This support is very gratefully acknowledged.

REFERENCES

[1] J. Bishop, H. McKay, D. Parrot, and J. Allan, *Review of international research literature regarding the effectiveness of auditory bird scaring techniques and potential alternatives*. Central Science Laboratories, 2003.

[2] I. Potamitis, S. Ntalampiras, O. Jahn, and K. Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Applied Acoustics*, vol. 80, no. 0, pp. 1 – 9, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003682X14000024

[3] L. Klein, R. Mino, M. Hovan, P. Antonik, and G. Genello, "Mmw radar for dedicated bird detection at airports and airfields," in *Radar Conference, 2004. EURAD. First European*, Oct 2004, pp. 157–160.

[4] C. Pornpanomchai, M. Homnan, N. Pramuksan, and W. Rakyindee, "Smart scarecrow," in *Measuring Technology and Mechatronics Automation (ICMTMA), 2011 Third International Conference on*, vol. 3, Jan 2011, pp. 294–297.

[5] C. Gonzalez, P. Melin, J. Castro, O. Castillo, and O. Mendoza, "Optimization of interval type-2 fuzzy systems for image edge detection," *Applied Soft Computing Journal*, vol. 47, pp. 631–643, 2016.

[6] C. Gonzalez, P. Melin, J. Castro, O. Mendoza, and O. Castillo, "An improved sobel edge detection method based on generalized type-2 fuzzy logic," *Soft Computing*, vol. 20, no. 2, pp. 773–784, 2016.

[7] S. Y. Na, D. Shin, J. H. Jung, and J. Y. Kim, "Protection of orchard from wild animals and birds using usn facilities," in *The 2nd International Conference on Computer and Automation Engineering (ICCAE), 2010*, vol. 3, Feb 2010, pp. 307–311.

[8] J. Jerabek, L. Zaplatilek, and M. Pola, "A proposal of radio ultrawideband systems for precision indoor localization," 2015, pp. 355–358.

[9] K. Mihreteab, M. Iwahashi, and M. Yamamoto, "Crow birds detection using hog and cs-lbp," in *Intelligent Signal Processing and Communications Systems (ISPACS), 2012 International Symposium on*, 2012, pp. 406–409.

[10] M. Shakeri and H. Zhang, "Real-time bird detection based on background subtraction," in *Intelligent Control and Automation (WCICA), 2012 10th World Congress on*, 2012, pp. 4507–4510.

[11] D. Moore, "Demonstration of bird species detection using an acoustic wireless sensor network," in *Local Computer Networks, 2008. LCN 2008. 33rd IEEE Conference on*, 2008, pp. 730–731.

[12] B. Diop, D. Diongue, and O. Thiare, "Bird sounds detection using normalized audio power sequences," in *2015 4th International Conference on Modeling and Simulation (MAS)*, 2015, pp. 43–48.

[13] V. Viswanathan, J. Makhoul, R. M. Schwartz, and A. Huggins, "Variable frame rate transmission: A review of methodology and application to narrow-band lpc speech coding," *Communications, IEEE Transactions on*, vol. 30, no. 4, pp. 674–686, Apr 1982.

[14] R. Sun, Y. Marye, and H.-A. Zhao, "Fft based automatic species identification improvement with 4-layer neural network," in *Communications and Information Technologies (ISCIT), 2013 13th International Symposium on*, Sept 2013, pp. 513–516.

[15] W. R. Gardner and B. Rao, "Theoretical analysis of the high-rate vector quantization of lpc parameters," *Speech and Audio Processing, IEEE Transactions on*, vol. 3, no. 5, pp. 367–381, Sep 1995.

[16] C. Rader and N. Brenner, "A new principle for fast fourier transformation," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 3, pp. 264–266, Jun 1976.

[17] J. Markel and A. Gray, *Linear Prediction of Speech*. Springer Berlin Heidelberg, 1976, iSBN: 978-3-642-66288-1.

[18] H. Hermansky, "Perceptual linear predictive (plp) analysis for speech recognition," *Journal of the Acoustical Society of America*, 1990.

[19] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 4, pp. 357–366, Aug 1980.

[20] A. McIlraith and H. Card, "Birdsong recognition with dsp and neural networks," in *WESCANEX 95. Communications, Power, and Computing. Conference Proceedings., IEEE*, vol. 2, May 1995, pp. 409–414 vol.2.

[21] A. Glowacz and W. Glowacz, "Sound recognition of dc machine with application of lpc and metrics," in *Human System Interactions, 2008 Conference on*, 2008, pp. 322–327.

[22] J. Durbin, "Efficient estimation of parameters in moving-average models," *Biometrika*, vol. 46, no. 1, pp. 306–316, 1959.

[23] H. Hua, J. Fan, and L. Wu, "Fuzzy decision-making and evaluation of environmental quality improvement for residential area," in *Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on*, vol. 6, 2009, pp. 222–226.

[24] P. Škrabánek and T. P. Runarsson, "Detection of grapes in natural environment using support vector machine classifier," in *Proceedings of the 21st International Conference on Soft Computing MENDEL 2015*. Brno, Czech Republic: Brno University of Technology, 23–25 June 2015, pp. 143–150.

[25] Y. Liu, K. Xu, J. Song, X. Liao, and Y. Zhao, "The design and implementation of knowledge processing and decision-making model based on multi-class in agricultural expert system," in *Information Science and Cloud Computing Companion (ISCC-C), 2013 International Conference on*, 2013, pp. 824–828.

[26] C. Resch, F. Pineda, and J.-J. Wang, "Automatic recognition and assignment of missile pieces in clutter," in *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, vol. 5, 1999, pp. 3177–3181 vol.5.

[27] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1999, iSBN: 0132733501.

[28] H. Nguyen, N. Prasad, and C. Walker, *A First Course in Fuzzy and Neural Control*. Chapman and Hall/CRC, 2003, iSBN: 1584882441.

[29] J. L. Blue and P. J. Grother, "Training feed-forward neural networks using conjugate gradients," vol. 1661, 1992, pp. 179–190.

[30] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," vol. 3, 1990, pp. 21–26.

[31] M. Kubat, *An Introduction to Machine Learning*. Springer International Publishing, 2015.

[32] E. Kiktova, M. Lojka, M. Pleva, J. Juhar, and A. Cizmar, "Comparison of different feature types for acoustic event detection system," in *Multimedia Communications, Services and Security*, ser. Communications in Computer and Information Science, A. Dziech and A. Czyzewski, Eds. Springer Berlin Heidelberg, 2013, vol. 368, pp. 288–297.

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

[34] P. Dolezel, M. Mariska, and I. Taufer, "Possibilities of feedforward multilayer neural network classifier as a detector of pest birds in vineyards," *International Journal of Engineering Research in Africa*, vol. 18, pp. 184–191, 2015.