Automatic Bird Species Identification for Large Number of Species

Marcelo T. Lopes, Lucas L. Gioppo, Thiago T. Higushi, Celso A. A. Kaestner
Federal University of Technology – Paraná
Curitiba, PR, Brazil
teiderlopes@gmail.com, lucaslongen@yahoo.com, thiago.higushi@gmail.com, celsokaestner@utfpr.edu.br

Carlos N. Silla Jr.
School of Computing, University of Kent at Canterbury
Canterbury, UK
cns2@kent.ac.uk

Alessandro L. Koerich
Pontifical Catholic University of Paraná / Federal University of Paraná
Curitiba, PR, Brazil
alekoe@computer.org

***Abstract*—In this paper we focus on the automatic identification of bird species from their audio recorded song. Bird monitoring is important to perform several tasks, such as to evaluate the quality of their living environment or to monitor dangerous situations to planes caused by birds near airports. We deal with the bird species identification problem using signal processing and machine learning techniques. First, features are extracted from the bird recorded songs using specific audio treatment; next the problem is performed according to a classical machine learning scenario, where a labeled database of previously known bird songs are employed to create a decision procedure that is used to predict the species of a new bird song. Experiments are conducted in a dataset of recorded songs of bird species which appear in a specific region. The experimental results compare the performance obtained in different situations, encompassing the complete audio signals, as recorded in the field, and short audio segments (pulses) obtained from the signals by a split procedure. The influence of the number of classes (bird species) in the identification accuracy is also evaluated.**

***Keywords*-signal processing; pattern recognition; machine learning; bird species identification**

I. Introduction

Bird species identification is a well-known problem to ornithologists, and is considered as a scientific task since antiquity. There are also practical reasons to monitor birds. In order to evaluate the quality of our living environment it is important to obtain reliable information about the population of wild animals. Birds are numerous and sensitive to environmental changes; also, and are easier to monitor than other species. Therefore, the use of automated methods for bird species identification is an effective way to evaluate the quantity and diversity of the birds which appear in a region [2], [3].

Another concern is the security of plane flights near airports. The Brazilian Center for Aeronautical Accident Investigation and Prevention (CENIPA) [6] reported that in two years a total of 1.321 aerial accidents involving bird collision with planes occurs in the Brazilian airspace, resulting in financial losses to the airline companies higher than US$ 3 millions. Furthermore, this type of accident is potentially dangerous: the collision of a vulture with a commercial plane is equivalent to an impact of seven tones [20]. With more information, authorities can do specific regulations to eliminate the problem.

The practical reasons previously mentioned justify the study of mechanisms for bird species identification. In this paper we focus on the automatic identification of bird species using signal processing and machine learning techniques. Until few time ago it was necessary to have a direct contact with a bird in order to determine its species. This situation has changed recently, with the use of automatic computational devices. One way to indirectly identify the species of a bird is using its recorded song. This task can be accomplished using signal processing [16] and machine learning techniques [21], [29]. This process includes:

1) the bird song recording in the field;
2) the use of audio preprocessing techniques to improve the signal quality, because these recordings usually occur in noise environments;
3) the extraction of features from the audio signal;
4) the use of this features in machine learning algorithms to produce a decision procedure that is able to predict the bird species.

In the classical supervised learning framework, a database of several recorded songs previously labeled by an expert is necessary to train the classifier. Several algorithms based on different paradigms can be used, such as probabilistic and instance-based classifiers, decision trees, neural networks and support vector machines [21]. The general framework is outlined in Figure 1.

The use of such an indirect way to monitor birds is in

978-0-7695-4589-9/11 $26.00 © 2011 IEEE
DOI 10.1109/ISM.2011.27

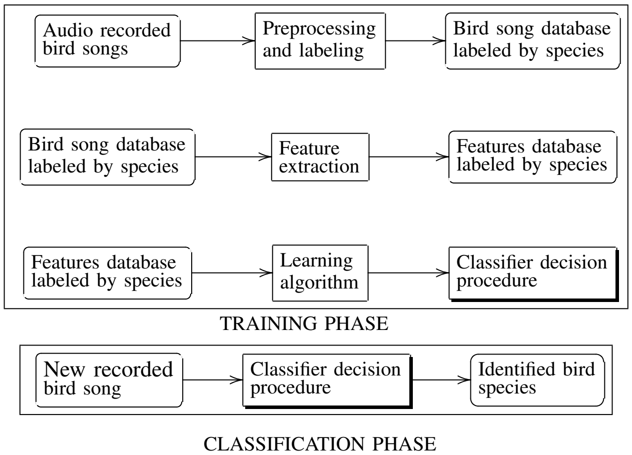

Fig. 1. The general classification framework

conformity with another contemporaneous concern: the welfare of the observed bird. The Canadian Council on Animal Care uses a scale – that goes from A (the least invasive) to E (the most invasive) – to measure the invasiveness (and subsequently pain) in any procedure that involves animals [4]. According to this scale, any indirect procedure has an invasiveness degree A.

As in most of the machine learning problems, the choice of the adequate feature set and classification algorithm are the main issues to be considered in order to obtain good classification performance [29]. In this paper we use the feature set produced by the MARSYAS framework [27], which is successfully employed in automated music genre classification problems [13], [22], [23], [27]. Furthermore, we consider two situations: the first one employs the bird songs as recorded in the field; the second one uses a preprocessing step that splits the recorded audio according to pulses, defined as short intervals with high amplitudes. We argue that pulses better characterize bird vocalization, and their use outperforms the use of the complete sound records [18]. In the two experimental scenarios we compare the performance of 5 classifiers with different paradigms in a dataset composed by bird sound recordings from 75 species. All the songs in this database belong to birds that appear in the Southern Atlantic Coast of South America.

This paper is organized as follows: Section II presents a formal view of the problem and summarizes related works; Section III describes the employed feature set, preprocessing steps and classification algorithms; Section IV describes the dataset and presents the experimental results obtained; finally, the conclusions of this work and future research directions are presented in the last section.

II. Bird Species Identification using Songs

As previously mentioned, the identification of bird species is a standard problem for ornithologists. With the use of autonomous devices it is possible to record a bird song directly in its natural environment. Acoustic communication is rich among birds, and it is one of the most employed way to detect them, even when it is difficult to see the bird itself. Therefore, the use of bird songs is considered as one of the most efficient ways to monitoring birds [3].

A. Problem definition

We define the automatic bird species identification problem as the task of finding the species of a specific bird from its recorded singing sounds. Bird sounds are usually classified as songs and calls. Songs are more melodious and are related to mating, and calls are employed as an alert signal or for other communications [5]. Calls usually correspond to very short and transient sounds; bird songs are longer and melodious, and are considered by experts to be ideal for bird species identification [5], [28]. Therefore, we consider only records of bird songs.

In our digital era, the analog signal is sampled, several times per second, and transformed by an analog-to-digital converter into a sequence of numeric values in a convenient scale. This numeric sequence represents the audio signal, and is employed in sound reproduction and for analytical processes [15].

In this context, if S denotes the audio signal, then $S =< s_1, \dots s_N >$ where each sequence element s_i stands for the sample obtained from the signal in the instant i, and where N is the total number of samples. The signal S contains a large amount of acoustic information, and several features can be extracted from it. Features are obtained directly from S (or from some part of it) by extraction functions. If χ is an extraction function and X_j is the feature domain, it is possible to obtain the feature vector $\bar{X} =< x_1, \dots x_D >$ from S, where each feature $x_j = \chi(S)$.

The bird species identification problem can be straightforward formalized as a pattern recognition problem [10]: given a bird song signal S as input, it is necessary to choose one class $\hat{b}$ from a finite set of bird species $\mathcal{B}$ that best represents the species of the bird that produced the song. If we employ a classical probabilistic framework, the problem can be formalized as follows. Given a set of evidences $\bar{X} = \chi(S)$ obtained from the input signal S, we must determine the class $\hat{b} \in \mathcal{B}$ with highest probability. That is:

$$\hat{b} = \arg\max_{b \in \mathcal{B}} P(b|\bar{X}) \tag{1}$$

where $P(b|\bar{X})$ is the *a posteriori* probability that the song belong to a bird of the species b given the evidences $\bar{X}$. If $P(\bar{X}|b)$ is the probability of occurrence of the feature vector $\bar{X}$ in the class b, $P(b)$ is the *a priori* probability of the bird species b and $P(\bar{X})$ is the probability of occurrence of the feature vector $\bar{X}$, then the previous equation can be rewritten using the Bayes' rule as:

$$\hat{b} = \arg\max_{b \in \mathcal{B}} \frac{P(\bar{X}|b).P(b)}{P(\bar{X})} \tag{2}$$

If a database of bird songs previously labeled by the corresponding species is available, then it is possible to estimate the first two probabilities in the last formula using frequencies; the last one ($P(\bar{X})$) is usually unknown, but if the we compute the likelihoods for the entire set of bird species, then $\sum_{b\in\mathcal{B}} P(b|\bar{X}) = 1$ and we obtain the desired probabilities for each $b \in \mathcal{B}$ by

$$P(b|\bar{X}) = \frac{P(\bar{X}|b).P(b)}{\sum_{b\in\mathcal{B}} P(\bar{X}|b).P(b)} \quad (3)$$

Since the denominator of Eq.3 is the same for all classes, the solution is given by the class

$$\hat{b} = \arg\max_{b\in\mathcal{B}} P(\bar{X}|b).P(b) \quad (4)$$

B. Related works

The automatic bird species identification problem has recently received the attention of the research community. Somervuo, Härmä and Fagerlund [25] develop signal processing techniques for the problem. They use a sinusoidal modeling and the Mel-Frequency Cepstral Coefficients (MFCCs). They test their proposal in 14 common North-European bird species, their best accuracy result is about 71.3%. Vilches et al [28] attack the bird species identification problem using some data mining algorithms: ID3, J4.8, and Naïve Bayes. They consider that the identification of distinctive features is crucial in resource constrained applications, and investigate dimensionality reduction in relation to classification accuracy. Using a database containing 154 songs from 3 bird species, their best result was 98.39% using J4.8 with a feature set obtained with the Sound Ruler audio processing tool. Fagerlund [11] employs a global decision tree with Support Vector Machines (SVM) classifiers in each node to separate two species. The author employs two feature sets: MFCC and a set of low level signal parameters. His best result was 98% obtained in a database with 8 bird species. Chou, Liu and Cai [7] propose an enhanced syllable segmentation method based on Rabiner and Sambur endpoint detection method. This method is combined with a MFCC feature vector to deal with two problems: syllable detection and bird song section recognition. They use songs from a commercial CD with bird calls and songs of 420 bird species, with recordings made in the field. The best obtained recognition rate result was 73.19% using a backpropagation neural network.

Lee, Han and Chuang [17] present a method for automatic classification of bird species that splits the original signal in syllable segments, considered as the basic recognition unit; then MFCCs are calculated, using GMM and Vector Quantization (VQ) to find the most appropriate number of GMM components and cluster number of VQ for each species. In the experiments they obtain the best classification accuracy of 84% for 28 bird species. Chou and Liu [8] use a wavelet transformation to transform sections of the bird songs. Then the first five order MFCCs are computed, and same order MFCC are aligned. They use a neural network classifier in a database with 420 bird species, achieving 73.41% for the recognition rate. Graciarena et al. [14] explore different modeling techniques to improve bird species classification from audio records. The authors use an unsupervised approach to obtain approximate note models from acoustic features. The note models are used to create a recognition system by leveraging a phone n-gram statistical model developed for speaker recognition applications. The approach is competitive when compared to GMM for the same acoustic features, considered as a baseline. They use 9 bird species, their best result presents an equal error rate of 16.5% in all considered songs. Chu and Blumstein [9] analyze temporal, spectral and structural characteristics of Robin songs and syllables. This elements are clustered by comparing a distance measure defined as the average of aligned linear prediction analysis on frame level differences. Syllable patterns are inferred from the clustering results and are used in a HMM Robin song detector. The system achieves a F-measure of 75.8% in its best result, using a database with 78.3 minutes of bird song recording.

Lopes at al. [18] present a comparison of the performance of 3 feature sets combined with a series of machine learning algorithms applied to the bird species identification problem. Experiments were conducted in order to evaluate various combinations of feature sets and classifiers in a database composed by 101 audio records from 3 bird species, which is similar to the one employed by Vilches [28]. The best obtained result indicates a F-measure of 99.7%, obtained using the audio records split into short intervals (pulses), using the MARSYAS feature set and a Multi Layer Perceptron (MLP) classifier.

From the state of the art it is clear that several papers have attacked the bird species identification problem with various preprocessing techniques and different machine learning algorithms. Up to now the conclusions are somewhat vague and difficult to compare, since there is no standards and the employed databases are specific and contain few species. This paper extends the work of Lopes et al. [18], using a larger database that contains bird songs recorded in a specific geographic region. The main objective of this paper is to evaluate if one should divide the bird songs into pulses or to use the full audio signal in classification, and what is the expected classification performance that can be achieved when dealing with a great number of bird species.

III. The Automatic Bird Species Identification Framework

A. Feature Set

In this work we employ the MARSYAS framework [19], that was already used in several audio applications. The MARSYAS feature set encompasses means and variances of timbral features, calculated in the intervals, for the spectral centroid, rolloff, flux, the time-domain zero crossings, including the 12 initial MFCCs in each case; the set includes 64 features. This feature set was employed for the first time in bird species identification by Lopes et al. [18]. In this paper the MARSYAS feature set performance was compared with the IOIHC [13] and the Sound Ruler [26] feature sets. Their results shown that the MARSYAS feature set performance overcomes the other two feature sets for most of the employed classifiers. For this reason, in this work we employ only the MARSYAS feature set.

B. Classification algorithms

In the experiments we use a set of classifiers based on different paradigms which include: the classical probabilistic Naïve Bayes algorithm; the instance-based k nearest neighbors (kNN) with $k = 3$; the decision tree classifier J4.8; a MLP neural network trained with the back-propagation momentum algorithm; and the SVM classifier, using the Platt's Sequential Minimization Algorithm (SMO) implementation, with two different kernel functions: polynomial and Pearson VII function-based universal kernel. These set of classifiers was chosen in order to represent various paradigms, and to make possible to find what family of algorithms are most suited for the focused problem.

IV. The Employed Database and Conducted Experiments

A. Database

We employ a new database composed if songs from bird species originated in a common geographic region: the Southern Atlantic Brazilian Coast. There are two main ecosystems related to this region: the Atlantic forest which occurs in the coastal region of the Atlantic Ocean; and the Araucária forest, composed by typical trees that are very characteristic from this region. In order to obtain bird sounds, we use the Xeno-Canto website [30]. The audio records in this database were obtained directly in real environments, without any filtering or similar preprocessing, and thus contain sounds from other birds and animals as well as environmental noise. The methodology employed to create the database was the following:

- Using the map search facility of the website, that uses the geographic locations where the bird sounds were recorded, we select bird species that have a record in a radius of 250 km (about 150 miles) near the city of Curitiba in Brazil, as indicated in Figure 2.
- From the species returned, we select 75 species with the highest frequencies; as in some cases the number of instances was still small, we completed the database using recordings of the same birds made in other regions.
- The songs were downloaded from the website by an automatic information extraction procedure, that creates a database with all the fields that appear in the Xeno-Canto queries. These fields include the scientific name of the bird species, the name of the recordist, the date, time and geographical location of the recording, the type of the recording sound (song or call), as well as the bird recorded song itself.

As previously mentioned in this paper we consider only bird songs. As result we obtain a database with the complete audio recordings of 1,619 song recordings of 73 bird species (two species have only calls and were discarded).

A second database was derived from the initial one: the original audio recordings were split according to pulses. As explained we define "pulse" as a short sound interval with high amplitudes. These segments seems to better characterize the bird vocalization, and their use improves the identification

Fig. 2. Geographic location of the bird sound recordings

performance. We use the Audacity audio processing tool [1] to split the bird songs into pulses. The database includes 8,226 pulses obtained from bird songs of the same 73 bird species. The splitting process is outlined in Figure 3, which presents the original sound recording for a bird of the species "Cercomacra-Tyrannina" (Dusty Antbird) and its corresponding pulses.

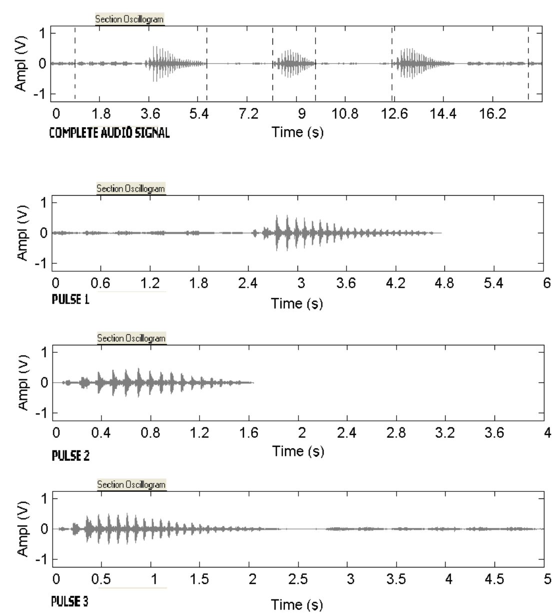

Fig. 3. Audio partition in song pulses

B. Experiments

The experiments were carried out in the two databases, with full audio recordings and with pulses. We use the MARSYAS framework to obtain the set of corresponding features. All the machine learning experiments reported in this section were carried out considering a 5-fold cross-validation procedure, that is, the presented results are obtained from 5 randomly

TABLE I
F-MEASURE ON THE FULL AUDIO RECORDS DATASET (%)

	Number of considered species				
Classifier	3	5	8	12	20
Naïve Bayes	61.5	50.7	27.0	25.3	25.4
kNN ($k = 3$)	61.4	53.4	41.5	33.1	33.0
J4.8	50.6	41.7	29.4	28.2	26.9
MLP	69.6	69.6	55.0	**48.8**	**47.4**
SMO (Polynomial)	**73.2**	**73.2**	**57.3**	47.2	46.4
SMO (Pearson)	67.6	59.5	51.8	42.3	42.7

TABLE II
F-MEASURE ON THE PULSES DATASET (%)

	Number of considered species				
Classifier	3	5	8	12	20
Naïve Bayes	45.9	32.9	27.4	24.8	17.6
kNN ($k = 3$)	93.4	88.1	83.8	81.9	77.3
J4.8	87.4	76.9	74.1	67.3	60.2
MLP	94.6	88.4	82.4	76.2	68.3
SMO (Polynomial)	85.5	75.0	72.2	65.8	59.6
SMO (Pearson)	**95.1**	**89.3**	**85.7**	**82.9**	**78.2**

TABLE III
F-MEASURE ON THE FULL AUDIO DATASET – RANDOM CLASSES (%)

	Number of considered species		
Classifier	3	5	8
Naïve Bayes	68.9	59.2	39.0
kNN ($k = 3$)	80.9	66.2	42.2
J4.8	60.4	51.8	36.9
MLP	**88.6**	71.4	**65.3**
SMO (Polynomial)	87.5	**75.2**	60.9
SMO (Pearson)	60.0	65.0	42.9

TABLE IV
F-MEASURE ON THE PULSES DATASET – RANDOM CLASSES (%)

	Number of considered species		
Classifier	3	5	8
Naïve Bayes	53.1	33.6	33.8
kNN ($k = 3$)	95.1	78.3	87.2
J4.8	86.4	69.9	72.8
MLP	**96.4**	**83.1**	87.4
SMO (Polynomial)	94.4	69.9	76.3
SMO (Pearson)	95.4	78.5	**89.7**

independent experiment repetitions. The experiments vary according to three dimensions:

1) the use of the full audio signal X the use of pulses;
2) the use of different classifiers: Naïve-Bayes, kNN with $k = 3$, J4.8, MLP, SMO-Polynomial and SMO-Pearson;
3) the number of classes; in this experiment we select the most frequent classes in the corresponding database.

The last dimension is used in order to evaluate the influence of the number of bird species (classes) in the classifier performance, which we consider crucial in real applications. We carried out experiments evaluating several number of classes, and the results obtained with 3, 5, 8, 12 and 20 classes are presented in the following.

Table I summarizes the results obtained in the first database, composed by full audio signals, with different classifiers and number of classes, selected as the most frequent ones in the corresponding database. The values in the Table represent the weighted average for the considered classes. The best results for each number of classes are emphasized. The best results for 3, 5 and 8 classes were obtained using the SMO with polynomial kernel, whereas for 12 and 20 classes the MLP achieve the top results.

Similarly, Table II presents the results obtained in the second database, with the same selection criteria. The best results for each number of classes are also emphasized. In all cases the SMO algorithm using Pearson VI kernel function provides the best results.

The analysis of the values presented in Tables I and II shows that the use of pulses instead of using the complete audio signal as recorded in field provides better classification results. This result corroborates with the ones obtained by Lopes at al. [18]. Our hypothesis is that the bird song files employed in the experiments present many silent intervals, where the bird sing itself is not present, but only the environmental noise. Also, the pulses encompass the most significant parts of the audio signal in terms of bird sound characteristics, which reflects in the features values extracted from them. We can also see clearly the drop in the classification performance when the number of classes increases. On the case of more complex algorithms, such as MLP and SMO this drop is lower, even so the correct classification is reduced in about 65% when going from 3 to 20 bird species. This situation shows that in real situations it is necessary to employ sophisticated algorithms and/or other information sources to assure better identification results.

A second experiment was made in order to evaluate the influence of the selected classes in identification. We select at random 3, 5 and 8 classes from the database, and the experiments were repeated. Tables III and IV show the corresponding results. For the complete audio signal the best results were achieved using MLP (for 3 and 8 species) and SMO-Polynomial (for 5 species). In the case of the pulses dataset the best results were also obtained by the MLP (for 3 and 5 classes) and by the SMO-Pearson (for 8 species). Here again, the use of pulses gives better results than the use of the complete audio signal. We also note that the results with 5 classes are inferior to the ones obtained with 8 classes. This situation emphasizes the obvious conclusion that classification performance is deeply influenced by the species involved in the classification: in real applications a previous study of the species which appear in the monitored ecosystem is fundamental to achieve good identification results.

A global analysis of the experiments described in this section can provide guidelines to be used in real applications. The most important conclusion is that to split the audio records in pulses in a simple and effective alternative to improve classification performance. Concerning to influence of the number of classes in classification, if we define a threshold of 80% for the F-measure as a good limit to be used in practical applications, we conclude that it is possible to identify up to 12 species using the proposed framework.

V. Conclusions and Future Work

This paper deals with the automatic bird species identification from bird audio recorded songs. We present a series of experiments conducted in a database composed by bird songs from 75 species which appear in the Southern Atlantic Coast of South America. Two audio datasets were considered: the first one is composed of bird songs as recorded in the field; the second one uses signal processing techniques to split the audio into pulses – short time intervals of the signal with high amplitudes. The experimental scenarios employ 5 classifiers with different paradigms to evaluate which one is more suited to the task. The number of species to be classified is also considered. We compare the performance of the algorithms for various sets of bird species selected from the database.

The experimental results show that the use of pulses is very important to improve the classification performance. Our explanation to this fact is that the acoustic information which appears in pulses contains less environmental noise and incorporates the most important features of the corresponding bird song. To obtain pulses from an audio signal is a very simple signal processing procedure, and can be easily incorporated in small computational devices, such as remote recorders.

Our best results using pulses were obtained with the MLP and SMO classifiers: for 3 classes, the obtained F-measure was 95.1% for the most frequent classes using SMO-Pearson and 96.4% for random selected classes using MLP; for 5 classes the corresponding F-measure values were 73.2% (with SMO-Pearson) and 83.1 (with MLP); and finally for 8 classes the values were 85.7% (with SMO-Pearson) and 89.7% (with SMO-Pearson). These numbers are in conformity with the ones obtained in most of the similar works, such as [9], [11], [14], [17], [18], [28]. The experimental results also show that using our simple approach we can obtain reasonable identification performance in problems with up to 12 bird species. In order to extend the identification capacity, we plan to apply hierarchical classification techniques [24] to the problem, using an ontology of bird species.

This study is also part of a broader project that encompasses hardware and software developments to monitor the bird species that appear in the urban area of the city of Curitiba (Brazil). The idea is to create a simple equipment, with some built-in processing capacity, to collect the acoustic information, that will be transmitted remotely to a central monitoring station. The collected data would allow ornithologists to monitor environmental conditions and to study specific bird species.

Acknowledgment

The authors would like to acknowledge the National Council for Scientific and Technological Development (CNPq) Fundação Araucária for the financial support.

References

[1] Audacity Web Site,<http://audacity.sourceforge.net/>, accessed in June 24th., 2011.

[2] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K-H. Tauchert and K-H. Frommolt, "Detecting Bird Songs in a Complex Acoustic Environment and Application to Bioacoustic Monitoring", *Patt. Recog. Letters*, Vol.31, No.12, pp.1524–1534, 2010.

[3] T.S. Brandes, "Automated Sound Recording and Analysis Techniques for Bird Surveys and Conservation", *Bird Cons. Int'l*, Vol.18, pp.163–173, 2008.

[4] Canadian Council on Animal Care Web Site, <http://www.ccac.ca/en/About_CCAC/About_CCAC_Main.htm >, accessed in June 24th., 2011.

[5] C.K. Catchpole and P.J.B. Slater, *Bird Songs: Biological Themes and Variations*, Cambridge University Press, 1995.

[6] CENIPA (Brazilian Center for Aeronautical Accident Investigation and Prevention), "The danger of the fauna for Brazilian aviation" (*in Portuguese*), < http://www.cenipa.aer.mil.br/cenipa/index.php/artigos-cenipa/118-o-perigo-da-fauna-na-aviacao-civil-brasileira >.

[7] C-H. Chou, P-H. Liu and B. Cai, "On the Studies of Syllable Segmentation and Improving MFCCs for Automatic Birdsong Recognition", *Asian Pacific Serv. Comp. Conf.*, Yilan, Taiwan, pp.745–750, December 2008.

[8] C-H. Chou and P-H. Liu, "Bird Species Recognition by Wavelet Transformation of a Section of Birdsong", *Symp. and Workshop Ubiq., Auton. Trusted Comput.*, Brisbane, Australia, pp.189–193, July 2009.

[9] W. Chu and D.T. Blumstein, "Noise Robust Bird Sonmg Detection using Syllable Pattern-Based Hidden Markov Models", *Int. Conf. on Acoust., Speech, Signal Process.*, Prague, Czech Republic, pp.345–348, May 2011.

[10] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification*, John Wiley and Sons, 2nd. Ed., 2001.

[11] S. Fagerlund, "Bird Species Recognition Using Support Vector Machines", *EUSASIP J. Adv. Signal Process.*, Vol.2007, pp.1–8, 2007.

[12] S. Garcia and F. Herrera, "An extension on "Statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons", *J. Mach. Learning Research*, Vol.9, pp.2677–2694, 2008.

[13] F. Gouyon, S. Dixon, E. Pampalk and G. Widmer, "Evaluating rhytmic descriptions for music genre classification", *Int. AES Conf. Virtual, Synth. Entert. Audio*, London, UK, 2004.

[14] M. Guaciarena, M. Delplanche, E. Shriberg and A. Stolcke, "Bird Species Recognition Combining Acoustic and Sequence Modeling", *Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, pp.341–344, May 2011.

[15] S. Hacker, *MP3: The Definitive Guide*, O'Reilly Publishers, 2000.

[16] B.P. Lathi, *Signal Processing and Linear Systems*, 2nd. ed. Oxford University Press, 2004.

[17] C-H. Lee, C-C Han and C-C. Chuang, "Automatic Classification of Bird Species from their Sounds using Two-Dimensional Cepstral Coefficients", *IEEE Trans. Audio, Speech, Lang. Process.*, Vol.16, No.8, pp.1541–1550, 2008.

[18] M.T. Lopes, C.N. Silla Jr., A.L. Koerich and C.A.A. Kaestner, "Feature Set Comparison for Automatic Bird Species Identification", in *IEEE Int. Conf. on Syst., Man, Cybern.*, Anchorage, Alaska, October 2011, to appear.

[19] Marsyas Web Site, <http://marsyas.info/>, accessed in June 24th., 2011.

[20] C.A.F. Mendonça, "The management of the aerial danger in Brazilian airports" (*in Portuguese*). *SIPAER Report*, Vol.1, No.1, November 2009.

[21] T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.

[22] C. N. Silla Jr., C.A.A. Kaestner and A. L. Koerich, "Automatic Music Genre Classification using Ensemble of Classifiers", *Proc. of the IEEE Int. Conf. on Syst., Man, Cybern.*, Montreal, Canada, pp.1687–1692, 2007.

[23] C. N. Silla Jr., A. L. Koerich and C. A. A. Kaestner, "A Machine Learning Approach to Automatic Music Genre Classification", *Journal of the Brazilian Computer Society*, Vol.14, No.3, pp.7–18, 2008.

[24] C.N. Silla Jr. and Alex A. Freitas, "A survey of hierarchical classification across different application domains", *Data Min. Knowl. Disc.*, Vol.90(1–2), pp.31–72, 2011.

[25] P. Somervuo, A. Härmä and S. Fagerlund, "Parametric Representations of Bird Sounds for Automatic Species Recognition", *IEEE Trans. Audio, Speech, Lang. Process.*, Vol.14, No.6, pp.2252–2263, November 2006.

[26] Sound Ruler Web Site, <http://soundruler.sourceforge.net/>, accessed in June 24th., 2011.

[27] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals", *IEEE Trans. Speech Audio Process.*, Vol.10, pp.293–302, 2002.

[28] E. Vilches, I.A. Escolbar, E.E. Vallejo and C.E. Taylor, "Data Mining Applied to Acoustic Bird Species Recognition", *IEEE Int. Conf. on Patt. Recog.*, Hong Kong, China, pp.400–403, 2006.

[29] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, 2005.

[30] Xeno-Canto Web Site, <http://xeno-canto.org/>, accessed in June 24th., 2011.