Bird classification algorithms: theory and experimental results

Conference Paper *in* Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on · June 2004

DOI: 10.1109/ICASSP.2004.1327104 · Source: IEEE Xplore

9 authors, including:

Chiman Kwan
Signal Processing, Inc.

Gang Mei
Intelligent Automation, Inc.

George Zhao
Intelligent Automation, Inc.

K. C. Ho
University of Missouri

BIRD CLASSIFICATION ALGORITHMS: THEORY AND EXPERIMENTAL RESULTS

C. Kwan, G. Mei, X. Zhao, Z. Ren, and R. Xu   V. Stanford*, C. Rochet*, and J. Aube*   K.C. Ho+

Intelligent Automation, Inc.
7519 Standish Place, Suite 200
Rockville, MD 20855, USA
ckwan@i-a-i.com

* NIST
Room A216, Technology Building (225)
Gaithersburg, MD 20899, USA

+ U. Missouri at Columbia
123 Engineering Building West
Columbia, MO65211, USA

Abstract

To minimize the number of birdstrikes, a common method is to use microphone arrays to monitor and identify dangerous birds near the airport or some critical locations in the airspace. However, it was recognized that the range of existing ground-based acoustic monitoring devices is only limited to a few hundred meters. Moreover, the bird classification performance in low signal-to-noise environments such as airports is not very satisfactory.

This paper summarizes the development of a high performance bird classification system using Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM). Experimental results verified the classification performance.

1. Introduction

Bird strikes cause more than 2 billion dollars of damage each year [1]. According to a FAA report [1], five species of birds are most dangerous, namely, chipping-sparrow, herring gull, Canada goose, mourning-dove and red-shouldered hawk. It would be ideal to develop a bird monitoring system with bird classification capability.

In the past year, IAI has developed a prototype system that has the following key components: 1) a circular microphone array with 64 mics; 2) a fast data acquisition system that can acquire bird sounds at 22 kHz sampling rate; 3) algorithms with Direction of Arrival (DOA) estimation, beamforming, and bird classification.

This paper focuses our results on the bird classification algorithms. Section 2 summarizes the HMM approach and Section 3 reports the GMM approach. Experimental results using GMM for bird classification is included in Section 4. Finally, concluding remarks are given in Section 5.

2. Bird Classification Using HMM

Figure 1 shows the architecture of the HMM based bird classification algorithm. There are two parts in the proposed configuration: a bird sound monitoring system and a bird recognition system. The monitoring system was described in [2]. The recognition part consists of Principal Component Analysis (PCA), Vector Quantization (VQ), and Hidden Markov Model (HMM). The PCA is mainly used for data dimension reduction and feature extraction. VQ provides a code sequence that can be used to characterize bird sounds, and the HMM is used for bird classification. The detailed description of these techniques is presented next.

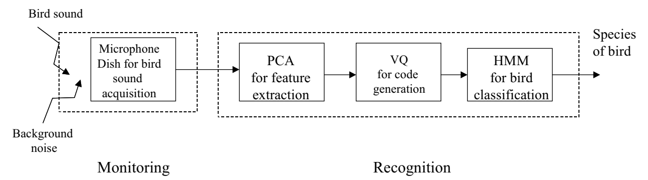


Fig. 1 HMM based bird classification system.

Preprocessing

Bird calls are usually stored in ".wav" or ".au" format and might be sampled at different rate. As a first step, the ".wav" and ".au" files were transformed into data files with the same sampling rate, say, 22050 Hz and normalized to the range of -1 to +1. Then the ceptral coefficients of the bird calls were used to detect and isolate call signal from the un-voiced period, followed by noise cancellation. Second, the isolated bird calls were blocked into frames of N samples, with adjacent frames being separated by M samples, as long as the frames were under the same call period. In other words, the first frame consists of the first N samples. The second frame begins M samples after the first frame, and overlaps it by $N-M$ samples, and so on until they covered the whole call period. This process is continued until all the sampled data is blocked into frames. Afterwards, a discrete Fourier transform is applied to each data frame and the power spectral density is computed. In order to get rid of some of the high frequency noise components, the obtained power spectral density is then passed to a low pass filter. More specifically, suppose there are totally L frames formed, the filter takes the following form:

$$x_i(k+1) = \alpha\, x_i(k) + (1-\alpha)\, u_i(k+1), \quad \text{for} \quad i = 1,\cdots,N, \quad k = 0,\cdots,L-1\,,$$

where $x_i(k+1)$ represents the i-th filtered data sample in frame $k+1$, $u_i(k+1)$ is the i-th data sampled in frame $k+1$ before filtering, and $0<\alpha<1$ is some design constant. Here we use $N=512,\ M=150,\ \alpha=0.9$.

PCA

Figure 2 best illustrates the key ideas of PCA [3].

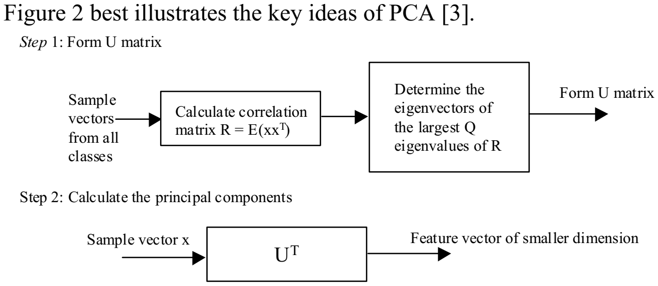


Fig. 2 Basic principle of PCA.

In our simulations, three largest eigenvalues of the correlation matrix are retained, and their corresponding eigenvectors are used to form the principal unit vectors for feature extraction. These three main eigenvalues usually represent more than 90% of the total energy of the input data.

VQ

It generates a sequence of codes to represent the bird call signature. The inputs to the VQ are the features from the PCA process. The code sequence generated by vector quantization captures the temporal characteristics of the bird calls. The particular VQ method we used is Learning Vector Quantization (LVQ), which is a supervised learning technique that uses class information to move the Voronoi vectors slightly, so as to improve the quality of the classifier decision regions.

The basic theory of LVQ is detailed in [3]. We improved the LVQ algorithm by also adjusting other weight vectors that are not the closest to the input vector. This additional adjustment avoids the situation that only one weight vector is moving all the time.

In our simulation, 20 Voronoi vectors are used to represent each fault class. Hence, the codebook has a size of 80.

HMM

To apply HMM for bird recognition, we need to address two important issues: training and state estimation. Fortunately, due to the past 30 years of work using HMM for speech recognition, these two issues have been solved. Efficient and fast algorithms are available.

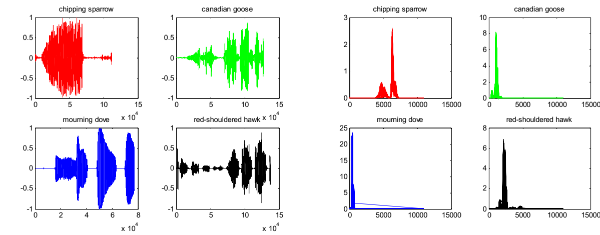
Training refers to the process of obtaining the HMM parameter set $\lambda = (A, B, \pi)$ [4]. In our proposed algorithm, raw sensor outputs first go through PCA for feature extraction and then the features go through a VQ to generate a sequence of codes. Finally, the code sequence goes into the HMM for training. One well-known approach is the Baum-Welch (or Expectation Maximization) method. In [4], this iterative procedure for choosing the HMM parameters was summarized.

For each type of bird, we use a HMM to model its acoustic behavior. Each HMM contains the characteristics of the state transition of an acoustic segment of the bird chirp. Each HMM is trained separately by using training data obtained from the microphone array. The data should contain various bird chirps.

Once the HMMs are trained, the model parameters are saved. During monitoring operations, the raw sensor data first goes through the feature extraction process using PCA. Then VQ is used to produce a sequence of codes. Finally, a parallel bank of HMMs is used to process the code sequences. Given the code sequence, the HMMs basically generate most probable state sequences that fits the given code sequences. The technique of estimating the HMM states is called the Viterbi algorithm, which is described in [4].

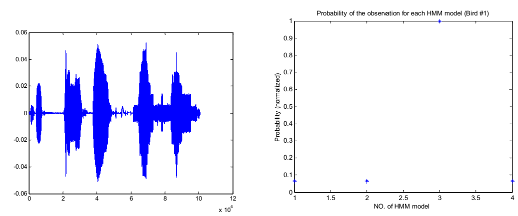
Simulation results

We used bird sounds from [5] [6]. One training set and the power spectrum are shown in Fig. 3. Figure 4 shows the results of testing. In this case, the HMM correctly identified the mourning dove. Table 1 summarizes other testing results for the other tests. It can be seen that although there are some misclassifications in some of the codes, the overall classification is quite good because the classification is performed by using the whole code sequence.



(a) Training data (b) Power spectral density

Fig. 3 Training data set 1.



(a) (b)

Fig. 4 (a) The testing data of mourning dove and (b) the decision of HMM. The decision sequence is 3-3-3-3-3-3.

Table 1 Bird classification results for 4 bird species.

Bird Species	Set number	Decision sequence	Overall probability
Chipping sparrow	#2	1-1-1-1-1-1	1.0000 0.0008 0.0121 0.0238
	#3	1-1-1	1.0000 0.0008 0.0121 0.0238
	#4	1-1-1-1	1.0000 0.0008 0.0121 0.0238
	#5	1-1-1-1	1.0000 0.0024 0.0324 0.0476
Canada goose	#2	2	0.0775 1.0000 0.0846 0.0789
	#3	2-2-2-2-2-2-2-2-2-2-2-2-2-2-2-2-2-2-4	0.1144 1.0000 0.1246 0.2214
	#4	2-2-2-2-2-2	0.0992 1.0000 0.1302 0.2080
Mourning dove	#2	3-3-3-3-3-3-3-3	0.0641 0.0641 1.0000 0.0641
	#3	3-3-3-3-3-3-3	0.0648 0.0649 1.0000 0.0649
	#4	3-3-3-3-3-3	0.0943 0.0939 1.0000 0.0932
	#5	3-3-3-3-3-3	0.0992 0.1052 1.0000 0.0982
Red-shouldered hawk	#2	4-4-4-4-4-1-1-4-4-4-4-4-4	0.3763 0.3331 0.1947 1.0000
	#3	4-4-4-4-4-4-4-4-4-4	0.1678 0.3986 0.1432 1.0000
	#4	4-4-4-4-4-4-4-4-4-4-4-4-4	0.1455 0.3347 0.1403 1.0000
	#5	4-4-4-4	0.1119 0.3926 0.1119 1.0000

3. Bird Classification Using GMM

GMM has been widely used for human speaker verification [7] [8]. The GMM based bird classification consists of two major steps: 1) preprocessing the extract features; 2) applying GMM models to classify different birds.

Preprocessing to extract features of birds

To identify the bird species, the algorithm is to first extract the feature vectors from the bird sound data, then match these feature vectors with Gaussian Mixture Models, each trained specifically for each bird class. The difference between the probabilities is compared to a pre-set threshold to decide if a given bird sound belongs to a specific bird class.

The feature extraction subsystem can be described by Fig. 5.

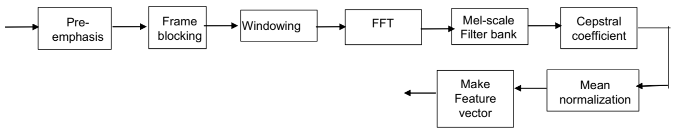


Fig. 5 Preprocessing steps in the feature extraction subsystem.

The purpose of feature extraction is to convert each frame of bird sound into a sequence of feature vectors. In our system, we use cepstral coefficients derived from a Mel-frequency filter bank to represent a short-term bird speech spectra. The digital bird sound data is first preprocessed (pre-emphasized, set to overlapped frames and windowed) and then Mel Frequency Cepstral Coefficient Analysis is applied. Typically feature extraction process compresses around 256 samples of bird sound data down to between 13 to 40 features.

Due to page limitation, the details of the feature extraction algorithm can be found in the Phase 1 final report [9].

GMM description

A Gaussian mixture density is a weighted sum of M component densities, given by the equation

$$p\left(\vec{x}\,\middle|\,\lambda\right) = \sum_{i=1}^{M} p_i b_i\left(\vec{x}\right)$$

where $\vec{x}$ is a D-dimensional random vector, $b_i\left(\vec{x}\right), i = 1,\ldots,M$, are the component densities and $p_i, i = 1,\ldots,M$, are the mixture weights. Each component density is a D-variant Gaussian function of the form

$$b_i\left(\vec{x}\right) = \frac{1}{\left(2\pi\right)^{D/2}\left|\Sigma_i\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\vec{x}-\vec{\mu}_i\right)' \Sigma_i^{-1}\left(\vec{x}-\vec{\mu}_i\right)\right\}$$

with mean vector $\vec{\mu}_i$ and covariance matrix Σ_i. The mixture weights satisfies the constraint that $\sum_{i=1}^{M} p_i = 1$. The mean vectors, covariance matrices and mixture weights from all component densities parameterize the complete Gaussian mixture density. These parameters are collectively represented by the notion

$$\lambda = \left\{p_i, \vec{\mu}_i, \Sigma_i\right\} \quad i = 1,\ldots,M.$$

For bird classification, each bird is represented by a model λ. The GMM can have several different forms depending on the choice of covariance matrices. The model can have one covariance matrix per Gaussian component (nodal covariance), one covariance matrix for all Gaussian components in a speaker model (grand covariance), or a single covariance matrix shared by all speaker models (global covariance). The covariance can also be full or diagonal.

GMM model parameter estimation

A free Matlab toolbox called Netlab was used to perform the GMM model estimation. The toolbox was developed by Ian T. Nabney at Aston University in the UK and it provides many useful Matlab functions for speech processing.

Bird classification algorithm

For bird classification, a group of S bird classes is represented by GMM's $\lambda_1, \lambda_2, \ldots, \lambda_S$. The objective is to find the bird model, which has the maximum a posteriori probability for a given observation sequence [8]. That is

$$\hat{S} = \arg \max_{1\le k\le S} \Pr\left(\lambda_k\middle|X\right) = \arg \max_{1\le k\le S} \frac{p\left(X\middle|\lambda_k\right)\Pr\left(\lambda_k\right)}{p\left(X\right)}$$

where the second equation is due to Bayes' rule. Assuming equally likely birds (i.e., $\Pr\left(\lambda_k\right) = 1/S$) and noting that $p\left(X\right)$ is the same for all bird models, the classification rule simplifies to $\hat{S} = \arg \max_{1\le k\le S} \Pr\left(X\middle|\lambda_k\right)$ using logarithms and the independence between observations, the bird identification system computes

$$\hat{S} = \arg \max_{1\le k\le S} \sum_{t=1}^{T} \log p\left(\vec{x}_t\middle|\lambda_k\right)$$

where

$$p\left(\vec{x}\middle|\lambda\right) = \sum_{i=1}^{M} p_i b_i\left(\vec{x}\right).$$

Simulation results

Four noisy signals were generated. Each noisy sample contains the desired signal, interferences and background noise. The noisy signals were passed through a beamformer to reduce the amount of background noise and interference. The beamformer output is then fed to the GMM for bird spices classification as shown in Figure 6.

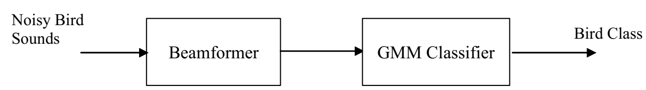


Fig. 6 Integration of the beamformer and the GMM bird classifier.

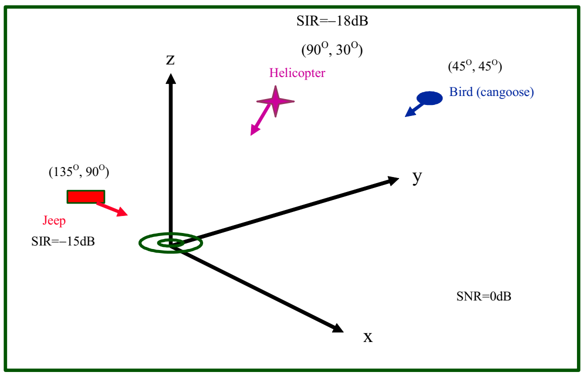


Fig. 7 Simulation Scenario 1.

Figure 7 shows a simulation scenario where the desired bird signal is Canada Goose. The two interferences are helicopter noise at –18 dB SIR and jeep noise at –15 dB SIR. The background noise is the Hoth noise at 0 dB SNR. Hoth noise, roughly speaking, is a lowpassed Gaussian noise with spectrum similar to voice. The beamforming and classification problem is more challenging if the background noise is Hoth. This is because the noise and signal spectra overlaps extensively in the frequency domain. As a matter of fact, the problem is easier if the noise is white.

The test samples are one of the training samples. Also shown below in Figure 8 is the received signal before and after beamforming, and the error between the true and the beamformed signal. Before beamforming, the noise and interference dominates. The bird source signal becomes apparent after beamforming.

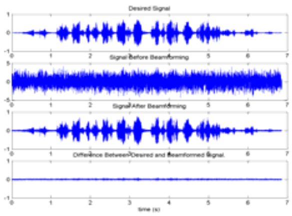


Fig. 8 Beamforming results from Scenario 1.

The classification results before and after beamforming are shown in Table 2. The more feature vectors we used the better the classification performance.

Table 2: Bird classification accuracy. Training and testing samples are the same. "fs" is the number of feature vectors used for classification.

	Percentage of Correct Classification					
	fs=1	fs=2	fs=3	fs=4	fs=5	fs=6
Before Beamforming	13.11%	12.27%	11.51%	12.29%	11.51%	10.77%
After Beamforming	100%	100%	100%	100%	100%	100%

Finally, we examine the classification accuracy when the input source signals are not from the training samples. The input test samples were the same as those when generating the results shown in Table 2. Table 3 gives the results before beamforming, and Table 4 is the results after beamforming. Before beamforming, the classification results are unsatisfactory as shown in Table 3. The classification results for Dove is high due to the fact that the classifier classifies the input as Dove most of the time, regardless of whether the actual bird is Dove or not. After beamforming, the classification results improve significantly. The identification results are all correct with fs=3, and they are very comparable to those shown in Table 2. The comparable results given in Table 2 and Table 4 indicate that the proposed circular array beamformer is very effective in reducing interference and background noise to improve the classification accuracy.

Table 3: Classification accuracy of four classes of birds. Test samples are different from the training samples, before beamforming

Bird Classes	Percentage of Correct Classification				
	fs=1	fs=2	fs=3	fs=4	fs=5
CanGoose	6.56%	2.20%	0.00%	0.00%	0.00%
Dove	81.72%	87.77%	88.04%	92.75%	90.91%
RSHawk	0.00%	0.00%	0.00%	0.00%	0.00%
Gull	28.98%	29.84%	28.35%	33.68%	28.95%

Table 4: Classification accuracy of four classes of birds. Test samples are different from the training samples, after beamforming

Bird Classes	Percentage of Correct Classification				
	fs=1	fs=2	fs=3	fs=4	fs=5
CanGoose	94.87%	94.74%	100%	100%	100%
Dove	98.86%	100%	100%	100%	100%
RSHawk	99.10%	100%	100%	100%	100%
Gull	97.40%	100%	100%	100%	100%

4. Experimental Results

In the bird monitoring experiment performed on June 16, 2003, we set up our microphone array in the parking lot, with one PC speaker playing bird sounds, another PC speaker playing aircraft noise.

We collected 6 sets of Canada goose sound, and 8 sets of chip sparrow sound for training the GMM model. Then we collected one set of Canada goose sound with and without aircraft noise, one set of chip sparrow sound with and without aircraft noise, for verification of the GMM model.

Table 5 is the result of bird classification. It can be seen that the GMM algorithm correctly identified the bird species.

Table 5: Experimental results of bird classification with beamforming

	Prob. as Canada goose	Prob. as chip sparrow	Decision
Canada goose test data without aircraft noise	**-89.1120**	-103.2840	Canada goose
Canada goose test data with aircraft noise	**-100.9485**	-149.3710	Canada goose
Chip sparrow test data without aircraft noise	-131.2033	**-92.0052**	Chip sparrow
Chip sparrow test data with aircraft noise	-110.9642	**-92.4776**	Chip sparrow

5. Conclusions

Two algorithms for bird classification have been presented. Based on our evaluations, GMM based algorithm yields better performance and is also suitable for real-time implementation.

Acknowledgment

This research was supported by the Air Force Office of Scientific Research under contract F49620-02-C-0044. Prior research supported by NIST under contract SB1341-02-W-1140 laid a solid foundation for this work.

References

[1] Wildlife strikes to civil aircraft in the United States. FAA report.

[2] C. Kwan et al., "An Automated Acoustic System to Monitor and Classify Birds," Bird Strikes Conference, August, 2003.

[3] S. Haykin, *Neural Networks*, Prentice Hall, 1989.

[4] L. Rabiner and B. Juang, "Fundamentals of Speech Recognition," Prentice Hall Signal Processing Series, 1993.

[5] www.math.sunysb.edu/~tony/birds/ and www.naturesongs.com/birds.html

[6] L. Elliott, D. Stokes, and L. Stokes, *Stokes Field Guided to Bird Songs: Eastern Region*, Time Warner, 1997.

[7] V. Stanford, *NIST Manual on Smart Flow System*, 1999.

[8] D. A. Reynolds and R. C. Rose, "Robust Text-Independent Speaker Verification Using Gaussian Mixture Speaker Models," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 1, 1995.

[9] C. Kwan et al., Phase 1 final report to the Air Force, July, 2003.