ON CLASSIFYING INSECTS FROM THEIR WING-BEAT: NEW RESULTS

Ilyas Potamitis±, Patrick Schäferǂ

±Technological Educational Institute of Crete, Dep. of Music Technology and Acoustics, potamitis@staff.teicrete.gr
ǂ *Zuse Institute Berlin* (ZIB), patrick.schaefer@zib.de

ABSTRACT

Insects variously affect many kinds of cultivations that are vital for rural economy, local heritage and environment: it is well known that insects pollinate a large number of plant species, while certain kinds of insects are pests that have a detrimental effect on cultivations. On top of the hazard list, mosquitoes can transmit serious diseases to humans and livestock. Pests can be controlled with aerial and ground bait pesticide sprays, the efficiency of which depends on knowing the time and location of insect infestations as early as possible. Automatic monitoring traps can enhance efficient monitoring of flying pests by identifying and counting targeted pests.

This work deals with novel advanced feature extraction and classification techniques as applied to the task of classifying insects from their wing-beat. It reports the most accurate results in the literature on two different datasets coming from a large number of flying insect species.

Index Terms— automatic insects classification, automatic monitoring of insect traps

1. INTRODUCTION

Producers set up traps in the field that lure and capture pests, in order to detect and count them. Manual inspection of traps is a procedure that is both costly and error prone. Inspection must be carried out manually as a repeated process, sometimes in areas that are not easily accessible. These traps can also, and often do accidentally capture other species of insects rendering the counting difficult especially for species that are morphologically similar. Pest surveillance, adaptable for other flying insect pests can provide pest information at regional and national scales, giving authorities a powerful tool to understand at a higher level the impacts and risks imposed by the presence of these pests.

A novel approach to wing-flapping recording devices has been recently announced [1]. The core idea behind these new sensors is to embed in insect cages or insect traps a device to record the fluctuations of light received by a photoreceptor as an insect flies through a laser beam and partially occludes light with its flapping [2-4]. The '*sonification*' of the baseband signal produced by the light fluctuations represents normal sound that is subsequently recorded using audio recorders and, therefore, in what follows it will be referred to as audio. This work focuses on feature extraction techniques and machine learning approaches as applied to the output of these sensors and not on the hardware per se.

Prior work in the field is reported in [3] where a large number of features are put to work along with some well-established classifiers. Another classification approach is described in [4] based on the not-widely used Complex Gaussian mixtures.

In this paper we put new work in context, by providing algorithmic details associated with two different datasets that have been made publicly available by the University of California at Riverside. The recorded insects are: *Aedes aegypti*, *Anopheles gambiae*, *Apis mellifera*, *Cotinis mutabilis, Culex quinquefasciatus*, *Culex tarsalis*, *Culex tarsalis*, *Culex stigmatosoma*, *Culex stigmatosoma*, *Drosophila melanogaster*, *Fungus gnat*, *Musca domestica and Psychodidae diptera*.

Our contributions are as follows:

(a) We compare 1-nearest neighbor classifier (1-NN) based approaches with more complex machine learning based classifiers. 1-NN perform best when considering only a few distinct insect species. When dealing with a larger number of insect species machine learning based classifiers perform best. (b) We capitalize on a new feature space, which was not considered in these earlier studies, namely the addition of the Klatt and delta-spectrum. (c) We present comments on other aspects of the algorithms applied other than the accuracy. (d) Finally, we apply details in tuning machine learning techniques that currently achieve the best performance reported in the literature [2] on the task of classifying insects from their wing-beat using these particular sensors.

2. THE SIGNAL OF A WING-FLAP

The datasets have been made publicly available by the University of California at Riverside. Dataset 1 (D1_10) [2] is composed of 5000 recordings equally split among 10 insect classes, namely: 01: *Aedes aegypti* male, 02: *Fruit flies* mixed sex, 03: *Culex quinquefasciatus* female, 04: *Culex tarsalis* female, 05: *Culex tarsalis* male, 06: *Culex stigmatosoma* female, 07: *Culex stigmatosoma* male, 08:

Aedes aegypti female, 09: *Aedes aegypti* male, 10: *Fungus gnats mixed sex.*

In order to answer various aspects regarding class separability of insect species based on their wing-flap, a part of D1 (the first five classes) formed the dataset D1_5.

Dataset 2 (D2_9) is composed of 18115 recordings with a highly unbalanced split among 9 insect classes (see [3] for details), namely: 01: *Aedes aegypti*, 02: *Anopheles gambiae*, 03: *Apis mellifera*, 04: *Cotinis mutabilis*, 05: *Culex quinquefasciatus*, 06: *Culex tarsalis*, 07: *Drosophila melanogaster*, 08: *Musca domestica*, 09: *Psychodidae diptera.*

All recordings in both datasets were sampled at 16 kHz. Fig.1 reports the mean duration of recorded audio events for both datasets. The length of each event is related to the time an insect takes to pass the field of the laser-photoreceptor setting. This takes less than 200ms. We notice that the standard deviation is quite high for all species indicating possibly different behavioral modes or different entrance angles. Datasets 1 and 2 differ on their mean indicating possibly different acquisition hardware settings.

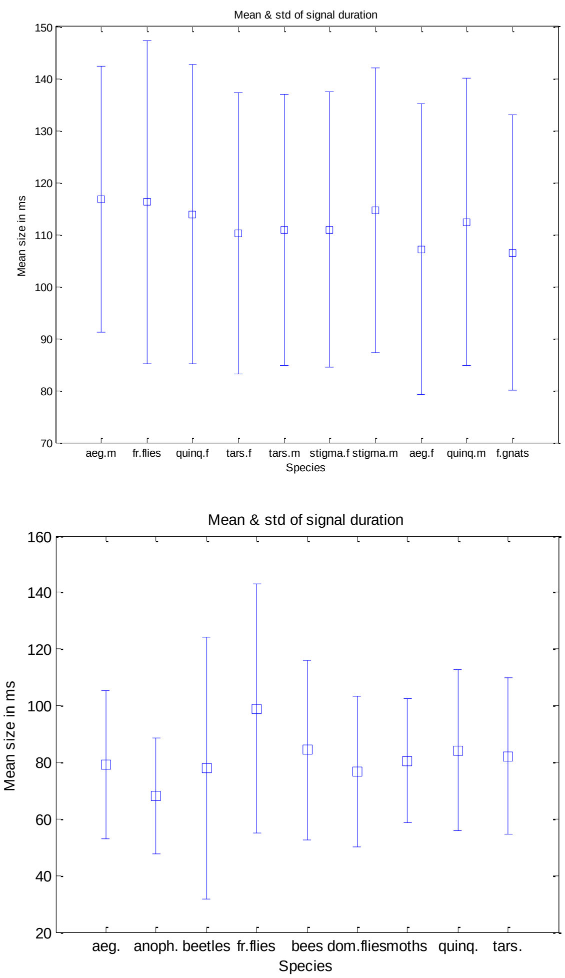

Fig. 1. Plots of the mean and ± standard deviation for the duration of an active event. Top: D1_10 dataset and bottom: D2_9 dataset.

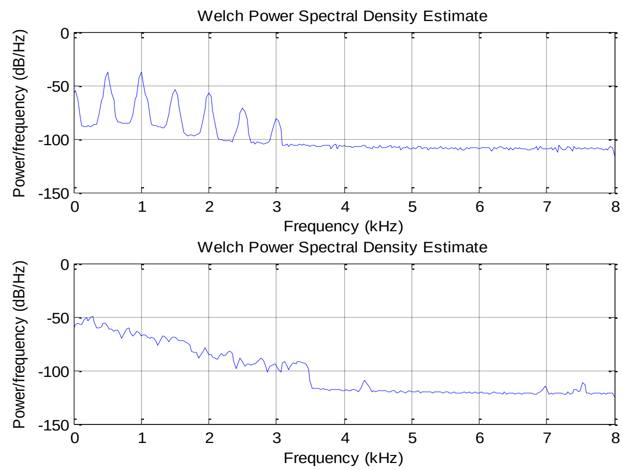

Fig. 2. Spectrograms of insects that are morphologically very different. Top: *Anopheles gambiae*, bottom: *Apis mellifera*.

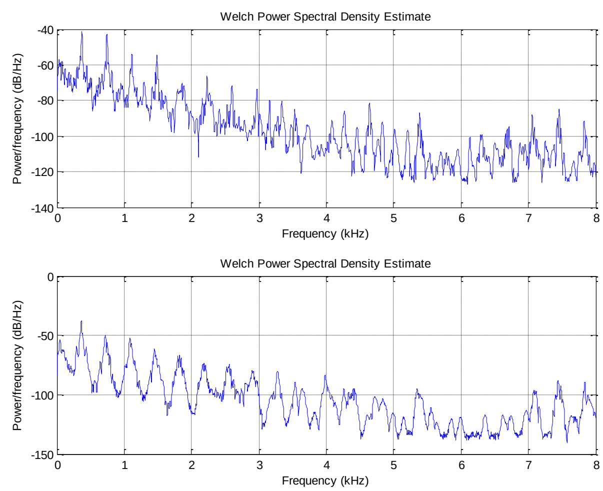

Fig. 3. Spectrograms of insects that are morphologically very similar: Top: *Culex quinquefasciatus* female, bottom: *Culex tarsalis* female.

Examples of the power spectrum of different cases are depicted in Figs. 2 & 3. In brief, the power spectrum of species that are morphologically very different differs significantly. Differences in the power spectrum between species that are morphologically very similar, like mosquitoes of the same sex but different species, are close to being acoustically and visually imperceptible. In such cases the smallest differences in the wings and thorax can in principle have an impact on the acoustic signature.

The classification results (see Section 5) demonstrate that the laser-photoreceptor hardware can produce an information stream that is adequate to resolve identity for different species up to a very high percentage, even for species that are quite similar.

3. FEATURES OF THE WING-FLAP SIGNAL

The audio feature vector comprises a summarization of the useful information (from the perspective of pattern classification) hidden in the sound signal. The ability to carry out species-independent detection and recognition lies in the selection of distinguishing acoustic features that remain relatively invariant regardless of the way the insect entered the laser scanned area. In [3] a variety of temporal and spectral features are evaluated as to their usefulness for classification. We have observed two things with regards to the features:

1) Some features, though theoretically can be associated quite naturally with the identity of an insect species, are in practice error prone in their calculation due to the short-time of the useful signal (e.g. the fundamental frequency-*f0* or the harmonics and their associated amplitudes). Therefore, any parameter than needs to be *estimated* from an original raw measurement and will be subsequently fed to a pattern classifier will be error prone especially for species that have similar size and morphology but are actually different.

2) The main source of misclassifications results almost completely from species that are very similar in size and morphology (e.g. *Aedes aegypti* male can be misclassified to *Culex tarsalis* male but never as *Apis mellifera* -the common bee) which is a much larger insect compared to mosquitoes. Note also that mosquitoes are dimorphic with females being larger than males. Therefore, acoustic discrimination of sex is not as difficult as it might seem on the first impression. In our experiments the most accurate approaches never confuse a *Culex quinquefasciatus* male with a *Culex quinquefasciatus* female. Therefore, temporal features such as the length of the passage time or zero-crossing rates do not offer discrimination advantage for morphologically similar classes that demonstrate highly overlapping temporal and spectral signal characteristics.

As a consequence of these observations we believe that the unprocessed spectrum and certain transformations of it (e.g. frequency pooling through a filter-bank) are a better choice than more sophisticated model-derived features (e.g. *f0*, harmonics, autoregressive features etc.). We therefore focus on this direction.

In the evaluation of all approaches we employed a common and simple approach to extract the active region of the recording. We found the peak and extracted a 2048 points centered window around the peak, roughly equal to 0.25s. That is, only the strongest part of the signal is extracted and subsequent analysis is based on a single audio frame.

4. DISTANCE MEASURES

Although classification accuracy is the crucial factor, other factors such as simplicity of implementation and computational cost are also of importance as the algorithms are meant to be embedded in stand-alone hardware devices. A highly accurate approach that would need a prohibitive amount of time and power in order to respond when embedded into hardware would be a poor choice against a suboptimal but simple and straightforward technique. In this work we examine two distinct categories of classification approaches namely: (a) the non-parametric ones that need no prior knowledge of the classes and (b) the parametric ones that make use of prior data. The non-parametric classifiers are based on a distance measure and 1-nearest neighbor classification (1-NN) and can operate directly on the data without any knowledge of the structure of the data or which species the data represents. They are simpler in their approach and easier to embed. The quality of the 1-NN classifiers crucially depends on the corpus of data to match against. A single outlier can significantly degenerate the classification quality. On average they achieved lower scores against model-based ones. Model-based classification approaches require a training phase so that they extract higher-level knowledge from labeled data, need a-priori the number of classes that they will classify, and are generally more complex than non-parametric ones. However, they are indisputably more accurate (with different levels of performance in each case). The a-priori knowledge of the number of classes in not a hard constraint for the type of application we aim for. That is the binary decision problem of the detection of a pest against all other classes that are not the target pest.

4.1. Non-parametric approaches

In what follows we analyze the basic concept behind each non-parametric approach. Each one is based on 1-NN classification. The label of an unlabeled recording Q is assigned to the label of the 1-NN S_i using a distance measure $D(Q, S_i)$. In the following we describe the distance measures used for 1-NN classification. Small symbols s_i represent feature vectors and large symbols S_i the recording.

4.1.1 Spectral dist.: Absolute norm of amplitude spectrum

Let q and s_i be the amplitudes of the Discrete Fourier Transform (DFT) of a recording with unknown label. The distance between Q and each member S_i of the set of recordings with known labels is calculated as: $D(Q,S_i) = \|q - s_i\|$.

4.1.2 Cepstral dist.: Square norm of weighted real cepstrum

Let q and s_i be the cepstrum (of a signal). The real cepstrum is the logarithm of the magnitude of the Fourier transform of a signal. The weighting is based on the autocorrelation function of the signal. That is:

$$\mathrm{s_i} = \mathrm{WR}\left\{\mathrm{S}_i\right\}\left|\mathrm{F}\left\{\log(|\mathrm{F}\{\mathrm{S}_i\}|^2)\right\}\right|$$

where S_i is the recording, $R\{S_i\}$ is the autocorrelation function and W is a linear window function that weights more initial cepstrum coefficients. The number of coefficients to keep was set to 100 by trial and error. The distance between Q and each member of the corpus with known labels $\mathrm{S_i}$ is then calculated as: $D(Q,S_i) = \|q - s_i\|^2$.

4.1.3 MFCC dist.: Square norm on cepstrum derived from the log of a Mel-scaled filter-bank

Let s_i be the power cepstrum (of a signal). s_i is first passed through a Mel-filter-bank represented by the matrix W that reduces within-band variation of the signal. Subsequently the log is applied to the filterbank output power. That is:

$$s_i = \left| \mathrm{DCT}\left\{ \log(W \left| F\{S_i\} \right|) \right\} \right|$$

where S_i is the recording and DCT denotes the discrete cosine transform. The distance between S_i and each member of the corpus with known labels S_i is then calculated as:

$$D(Q, S_i) = \left\| q - s_i \right\|^2.$$

4.1.4 Klatt dist.: Weighted slope distance metric

The signal is critically bandpass filtered (Bark scale) and the output power is found. Finally the first order difference across frequency of the log output power is derived. The so-called Klatt spectrum is associated with the formants area. It was not found to be the best metric but it achieves a descent score only by using the difference of the spectrum across frequency thus indicating that the difference spectrum can be used as an extra source of information appended to the feature set of the model-based classification approaches.

4.1.5 Symbolic Fourier Approximation (SFA) distance

We built feature vectors based on SFA [5] which was presented in the context of time series similarity search. It is a discrete representation of time series. The SFA transformation of a time series results in a sequence of discrete values. SFA consists of two operations:
1. Low pass filtering using Discrete Fourier Transform (DFT) and, 2. Quantization using a quantization technique called Multiple Coefficient Binning (MCB).

We extracted sliding windows from the audio recordings. The DFT converts each sliding window from the time domain to the frequency domain. Low pass filtering is applied to cut the frequency range to 0Hz and 7200Hz. The aim of quantization is to allow for a fuzzy matching of two audio recordings in the presence of noise (Observation 1). The number of quantization intervals for each DFT coefficient was fixed to 22. Matching two SFA feature vectors is then based on a method called shotgun analysis from bioinformatics. For the sake of brevity, we omit the exact details of this matching (see [5] for details).

5. RESULTS & DISCUSSION

We first evaluated the non-parametric approaches. We used the official D1_5 train/test split with 500 and 4500 recordings each. D1_10 train consists of 5000 recordings.

The simplest approach among all tested is the one based on the spectral absolute norm distance between two recordings. If there was a prize for simplicity this approach would score first rank as it could be realized with just one line of code. It does not achieve the best score but is at close distance from far more complicated distance-based approaches. The Cepstral distance approach is also simple with a direct implementation but shows increased sensitivity when enlarging the number of classes as shown in Table 2. A distinct improvement is observed in the MFCC square distance approach. Although it is a Cepstral distance approach it reduces the within-species variability by employing a Mel-scaled filterbank. This is also observed in human speech where the MFCC signals are widely accepted [6]. The Klatt distance, though it has quite an acceptable performance on D1_5 cannot handle adequately the many classes case of D1_10. One should note that D1_10 is the hardest dataset of all not only because it has 10 classes but also because many species are morphologically very close (e.g. mosquitos of different species but same sex) and this results in a very similar audio fingerprint. The ‘SVM on MFCC’ is a Support Vector Machines (SVM) classifier as applied to MFCCs.

Method	**Train**	**Test**	**Training**
SVM on MFCC	99.8	93.29	YES
MFCC’s distance	93.2	93.69	NO
Cepstral distance	91	87.89	NO
Klatt distance	92.6	91.73	NO
Complex GMM	100	89.71	YES
SFA distance	96.8	**94**	*tuning*
Spectral distance	88.2	91.71	NO

Table 1: D1_5 corpus 1-nearest neighbor classification.

Method	**Train**	**Training**
MFCC’s distance	68.12	NO
Cepstral distance	60.46	NO
Klatt distance	63.7	NO
SFA distance	**68.26**	*tuning*
Spectral distance	67.68	NO

Table 2: D1_10 corpus 1-nearest neighbor classification.

The SFA distance measure is not a completely non-parametric procedure as it needs to fine tune 3 parameters (sliding window-length, quantization intervals and low pass filter size) on the training data. It proves one of the top scoring approaches in the D1_5 and retains the best score over other distance based approaches in the more difficult D1_10 dataset. It suffers though in the case of increased number of classes like all non-parametric approaches.

The Complex GMM (CGMM) that is described in detail in [4] is also a parametric approach that requires a training phase on a labelled dataset. CGMM’s score favorably but are computationally intensive to train.

The joint examination of D1_5 in Table 1 and D1_10 in Table 2 sheds some light into the question if non-parametric approaches can score favorably against the more computationally intensive parametric ones. The results prove that the 1-NN classifier based non-parametric approaches drop significantly in performance as the number of classes and no.of.recs within classes increase. To become a solution in the most complex scenarios of many species classification careful selection of labeled samples for the 1-NN classifier would be necessary. This might be a direction of future work. 1-NN classification is sufficient when the number of species attracted to a specific trap is small and the classes are not spectrally very similar (like in the case of the binary decision problem). Once we reached this conclusion any further examination of datasets composed of a large number of insects is retained only for the parametric approaches. We subsequently focus on parametric classifiers that can deal with high-dimensional feature sets and we append to the dataset and spectrum-originating transformation that could prove useful for our task.

5.1. Model based approaches

Once we have fixed our approach to rely exclusively on the spectrum and its transformations we proceed to employing well-established machine learning techniques that are capable of dealing with high-dimensional datasets. These datasets result form the concatenation of different feature sets. Support Vector Machines (SVM), Random Forests (RF) and Extra Trees as well as Gradient Boosting Classifiers (GBC) are known to be able to handle efficiently high dimensional feature sets [7]. The Complex GMM make use only of the power spectrum as this is an inherent assumption of this algorithm and cannot handle any other feature set (see [4] for details). The Delta refers to the first order difference across frequency. The common feature set in Table 3 and Table 4 is a column stack concatenation of various feature sets having as a basis the amplitude spectrum: S = [S1|S2|S3|S4] totaling 264 features, where:
S1: MFCC (51 features), S2: Delta MFCC (50 features), S3: Power spectrum filter-bank (60 features), S4: Delta Power spectrum filter-bank (59 features), S5: Klatt spectrum (44 features)
The validation on D1_10 is based on 10-fold cross-validation due to non-availability of a test set. To make our results comparable to [3], we use their hold-out validation scheme (33% is retained for training and 67% for testing) on D2_9. 10-fold cross-validation is used on the training data.

Method	10-fold	Training
SVM	77.66	YES
Complex GMM*	74.99	YES
Random Forest	76.82	YES
Extra Trees	77.42	YES
GBC	79.94	YES

Table 3: D1_10 corpus 10-fold cross-validation scores.
* power spectrum only, validated on a 80-20 hold-out scheme.

Method	10-fold	Hold-out
SVM	88.69	87.33
Random Forest	87.36	86.96
Extra Trees	88.68	87.46
GBC	90.75	**89.24**

Table 4: D2_9 corpus 10-fold cross-validation scores on the train set and on the hold-out set.

The splitting of the training database to train and test is provided by [3]. All classifiers are tuned using greed-search on the training corpus. The large GBC of 3000 estimators demonstrates the best performance in all our experiments.

6. CONCLUSIONS

We present the best results so far in the literature on the subject on classifying insects based on their wing-flap for small and large number of distinct species. For a few species 1-NN classification in combination with SFA scores best reaching 94% on the official test/train split. For large numbers GBC scored best reaching 90.75% on 10-fold cross validation and 89.24% on a strict hold-out set.

ACKNOWLDGEMENTS

The first dataset was provided by Prof. Eamonn Keogh from the University of California at Riverside. The second dataset was made available here: http://sites.labic.icmc.usp.br/dfs/ICMLA2013/ (date last viewed 10/2/2014). The code used in all comparative experiments was kindly provided by the original authors of the code: SVM on MFCC features by Dominik Schnitzer OFAI, square norm on Mel-cepstrum by Dan Ellis (no LDA applied), absolute norm on spectrum by Antonio Deusany de Carvalho. Complex GMMs are described in [4] and the SFA approach in [5]. This work is funded by the FP7-SME-2013 EC project ENTOMATIC – 605073.

6. REFERENCES

[1] http://www.cs.ucr.edu/~eamonn/CE/ (d.l.v 10/2/2014)
[2] G. Batista, E. Keogh, A, Mafra- Neto, E. Rowton. Sensors and Software to allow Computational Entomology, an Emerging Application of Data Mining. SIGKDD 2011
[3] Silva et al., Applying Machine Learning and Audio Analysis Techniques to Insect Rec. in Intelligent Traps, ICMLA 2013.
[4] Potamitis I., Classifying insects on the fly, Ecological Informatics, http://dx.doi.org/ 10.1016/ j.ecoinf.2013.11.005, 2013.
[5] Schäfer P., Högqvist M., SFA: a symbolic Fourier approximation and index for similarity search in high dimensional datasets. EDBT Proceedings of the 15th International Conference on Extending Database Technology, pp. 516-527, 2012.
[6] X. Huang, A. Acero, H. Hon, "Spoken Language Processing," Prentice Hall, 1-1008, 2001.
[7] Pedregosa et al., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12: 2825-2830, 2011.