Interactive environmental sensing: Signal and image processing challenges

Allen, M. , Graham, E. , Ahmadian, S. , Ko, T. , Yuen, E. , Girod, L., Hamilton, M. and Estrin, D.

INTERACTIVE ENVIRONMENTAL SENSING: SIGNAL AND IMAGE PROCESSING CHALLENGES

Michael P. Allen, Eric Graham, Shaun Ahmadian, Teresa Ko, Eric Yuen, Lewis Girod, Michael Hamilton, Deborah Estrin

Center for Embedded Networked Sensing
University of California, Los Angeles

ABSTRACT

Networked embedded acoustic sensors and imagers allow scientists to observe biological and environmental phenomena at high sampling rates and multiple scales. Such sampling can create large data sets that often require some form of automated processing to extract useful information. However, to guarantee the accuracy of the data, the scientist must be included in the processing, rather than treating it as a black box, an approach we call *interactive environmental sensing*. In this paper we describe the challenges of such an approach and motivate it with several examples from bioacoustics, plant phenology and avian biology.

Index Terms— Image processing, Acoustic arrays, Environmental factors, Interactive systems, Automation

1. INTRODUCTION

Environmental sensing has long been a strong motivator for embedded systems research. The capabilities of networked embedded systems to increase sampling density and coverage have enabled scientists to study phenomena at scales previously impossible [1]. Acoustic sensors and imagers are especially important for capturing biological phenomena but with high frequency sampling can quickly create very large data sets. Although using these sensors can transform the observational capacity, traditional manual processing then becomes impractical. Often, the only way to analyse these data sets is to reduce them by automated means.

To answer scientific questions using the data collected by embedded networked sensing, scientists must have confidence in the reliability of the sensors and in the quality of the data. At a higher level, confidence is also required in the accuracy of automated detectors and classifiers used to reduce large data sets. Although automation is desirable in well-characterized processes, the ecological events being examined by environmental scientists may not be clearly defined or be so complex that events are difficult to automatically characterize a priori – they require iteration.

Michael P. Allen is a PhD student at Coventry University and Lewis Girod is a post-doctoral researcher at Massachusetts Institute of Technology.

The concept of interactive environmental sensing implicates the scientist as a necessary part of the process which is used to reduce the data, as well as enabling interaction when anomalous or unexpected data presents itself. This approach has advantages – first, the scientist can be assured that the data gathered is of the expected quality, and second that the automatic data reduction and analysis techniques are accurate.

This paper motivates and describes signal processing challenges presented by interactive environmental sensing. We discuss applications and challenges from three examples in two areas of signal processing: acoustic and image processing. Our examples are drawn from bioacoustics, plant phenology and avian biology. We describe the general problems facing these types of interactive sensing applications and subsequently discuss future challenges.

2. ACOUSTIC ARRAY

In bioacoustics research, being able to detect, classify and localize animal and bird vocalizations is an important part of understanding behavior. Traditional approaches involve either manual observation in the field by the scientist, or deploying a wired array of microphones over the area of interest to record data for offline analysis.

We have developed two generations of wireless acoustic monitoring boxes, designed for rapid, attended deployment and in-field processing [2]. These platforms feature a four-microphone sub-array, arranged in a tetrahedral configuration, enabling the use of techniques requiring highly coherent signals (such as beam forming) per-node. A network of acoustic boxes can time synchronize and self-localize to high accuracy using acoustic time of flight and direction of arrival techniques [2]. Being able to process vocalizations on-line allows the scientist to get an idea of the quality of the data that is being gathered, as well as enabling reconfiguration, to react to unforeseen events such as phenomena moving out of the area covered by the network. These types of interaction are not possible by analyzing an off-line data set.

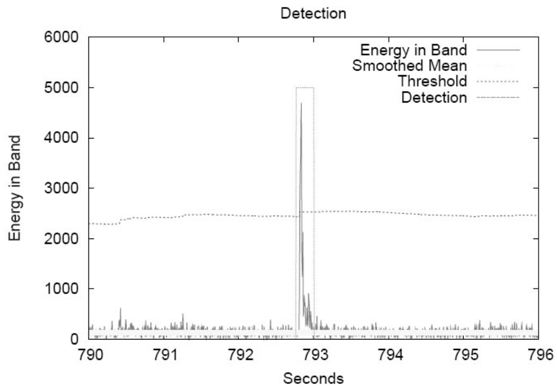

Fig. 1. Example of the detection of an event of interest against the estimated noise floor.

2.1. Event Detection

An approach is taken using an on-line event detector to discriminate events of interest in a continuous stream of audio based on energy in specific frequency bands. The detector assumes a Gaussian ambient noise model, adaptively estimating the noise floor using two Exponentially Weighted Moving Average (EWMA) filters for mean and variance. Field tests have shown good performance and suitability for on-line event detection of the vocalization of marmots [3].

The event detector is optimized by processing one channel of audio, decimating to 24KHz and processing only 1/4 of the samples taken (a typical marmot vocalization has a length of 0.04 seconds). The energy of the samples is determined in the frequency domain, taking the magnitude of the sum of pre-determined frequency bins of interest. This energy feeds in to the detector, and depending on its value is determined to be either noise or an event of interest. Figure 1 shows an example of a detection in the band of interest compared to the estimated noise floor.

This type of detector is ideal to run on-line on an embedded platform, and is immediately applicable to different species vocalizations, by tuning specific noise threshold and frequency interest parameters. These pre-determined characteristics would be observed by the scientist in the field, and used to adjust the detector.

2.2. Localization

Localizing an event of interest over a distributed network of acoustic boxes is a multi-step process of sub-array processing, event clustering and fusion. Thanks to the close microphone proximity on each sub-array, the Approximated Maximum Likelihood algorithm [3] can be utilized, which estimates the likelihood of direction of arrival (DoA) at each possible degree (or higher). This approach presents a trade-off between sub-array size and spatial aliasing which is comparable to the Nyquist limit; energy in wavelength frequencies lower than two times the microphone spacing will be aliased (see [3] for a detailed discussion).

Based on detection time, events must be clustered together such that they represent the same animal vocalization, and then suitably combined to form an overall estimate. This can be naively done by creating a likelihood map, where each point on the map represents the combination of the direction of arrival likelihoods of all nodes that detected the event [3].

3. PLANT PHENOLOGY

In plant phenology, yearly patterns in timing of bud burst, flower bloom and the numbers and sizes of leaves are important indicators of environmental conditions. Plant phenology has been identified as a crucial contributor to global change research [4]. Asynchrony of phenological events may disrupt plant and animal communities and be a signal of significant environmental change [5].

Collecting phenological data manually is time-consuming and labor-intensive and thus much basic ground-based phenological information is lacking [6]. The use of actuated imagers as biological sensors is therefore ideal for wider phenological studies, allowing a higher frequency of observations, especially over large areas; Imagers can also be left in remote locations which are troublesome for the scientist to repeatedly access.

3.1. Leaf and Flower Detection

At James Reserve (Idyllwild, CA), high-resolution, pan, tilt, zoom (PTZ) controlled cameras mounted on fixed towers make daily scans of their surroundings, aiming to capture plant phenological events (gathering about 1,200 images/day). The cameras must be zoomed in closely when scanning, due to the small size of some of the plants/flowers. Manually inspecting the resulting image streams for changes in leaves or flowers is an unreasonable task, so automated detection is desirable. Unfortunately while it may be easy to detect the presence of a well-developed flower using color characteristics, some flowers and leaves have weaker color characteristics and can be mostly missed by automated detectors.

The goal is to reduce the images in the data stream to only those which are candidates for containing phenological events. The scientist can then visually verify whether a flower is present in the candidate image, for instance. After this, the scientist can inspect images captured on previous or subsequent days at the same PTZ coordinates to find the exact timing of important events, such as bud burst or senescence.

Flowers and leaves being monitored vary in size, shape and color which affect how easy they are to detect. This is compounded by changes in lighting conditions from day to day. The easiest types of plant phenological events to automatically detect are the presence of numerous (or large), leaves or flowers that are in high contrast to the background.

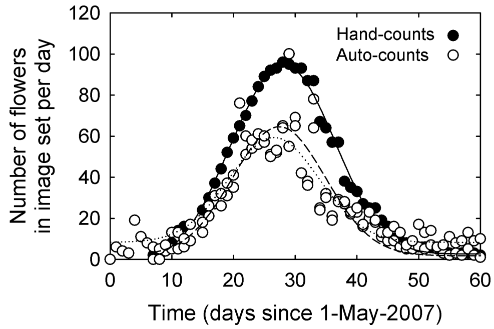

Fig. 2. Manual vs automated counting of blooming wall flower, *Erysimum capitatum*. The automated prediction of blooming events closely matches the manually counted one.

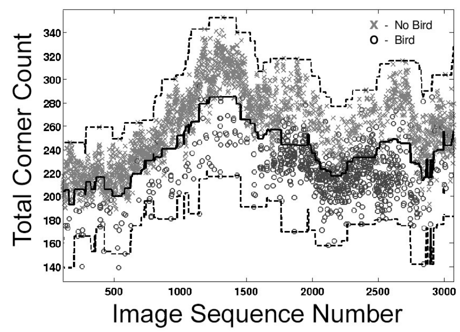

Fig. 3. An indication of the adaptation of the threshold indicating bird presence/absence, based on corner counting.

In these cases, using the Hue, Saturation, Lightness (HSL) color space is preferable to RGB (in which all three components change as light levels change). Hue represents a continuous variable for measuring color that is independent of lightness and saturation - thus, if the ambient light changes, hue will not change and the salient feature will still be detected.

3.2. Tools for Interaction

To provide the interaction with the data set that the scientist requires, a visualization tool has been implemented. The tool is a color filter applied to selected images. By setting the min/max values of individual color components, the tool provides a quick way to mask areas of an image. After a filter has been optimized, it can then be applied automatically to the entire image set, producing a reduced set of filtered images.

Other integrated tools are used to help improve contrast and reduce noise in the image (constrast stretching, histogram equalization, Gaussian blur). A rough flower counter has also been implemented (see Figure 2). For data quality assurance, the scientist can easily tag images, allowing the performance of the automated system to be measured against a ground truth.

4. AVIAN BIOLOGY

Avian biologists investigate trends and differences in behavior that affect reproductive success (and the effects of microclimate variability on this). Traditionally, avian biologists manually inspect nesting locations and visually log data about the state of the nest. The number of observations is increased by deploying imaging sensors (sampling every 15 minutes) in nest boxes over a given area. The avian biologist can then interactively use tools for nest box processing to target visual investigation of potentially anomalous or unforseen events.

4.1. Processing Goals and Challenges

The aim in processing the stream of images gathered at a nest box is to determine a bird's presence and absence, count number of eggs, and determine the precise timings of transitions among stages of the nesting cyle. The scientist is also interested in observing hatching, feeding and parental care behaviour.

Using imagers in fixed positions in the nest box is advantageous because the size of eggs can be characterized, and the area covered by the camera's field of view is constant. Also, because the image stream represents the nest over a time period, each image does not have to be processed as if it were not unrelated to the others.

However, nest box images can be influenced by variable lighting quality depending on night and day. Therefore, each nest box is lit using Infrared (IR), providing a consistent (non-disturbing) lighting source - this requires that all images captured must be grayscale. Sensor placement and sensor sensitivity affect the similarity of images across different boxes, and different species may build nests in different ways. This can limit the general applicability of any image processing techniques applied.

4.2. Bird and Egg Detection Techniques

Even though the data stream is temporal in nature, a simple approach (such as frame differencing) will not work given the relatively low sampling rate. Instead, birds are detected by taking advantage of interest points. In this case, corners (areas in the image where gradients are large in two directions) are used as interest points. These points can be found using a Harris-Stephens detector [7]. Since a bird's feathers have a smooth, approximately homogenous appearance, they will have a lower density of interest points than a more textured region, such as the nest. Bird presence can then be determined by taking the midpoint between minimum and maximum interest points over a four day window of images (around 200). This midpoint will change adaptively, and is used to decide whether a bird is present or not, as shown in Figure 3.

For images which have been determined not to contain birds, eggs are counted. This problem is approached not by trying to count the number of eggs in a single image (which can be difficult, even for a human), but by counting the number of eggs in the nest box over time. Each image is searched for blobs of interest, using a Scale Invariant Feature Transform (SIFT) detector [8]. Blobs are areas exhibiting maximal response to a Laplacian of Gaussian filter (LoG); the output of the SIFT detector gives a scale for the blob region. The mean and variance of intensity around this location is found to be characteristic of eggs and are used to discriminate between eggs and other egg-like objects. Using this output as input to a Hidden Markov Model (HMM) leverages the temporal constraints of the image stream, and accounts for the underlying statistic of occluded eggs and the existence of egg-like objects, to deduce a final egg count and transition in nesting stages.

5. CONCLUSION

In this paper, we have described interactive environmental monitoring and the signal processing challenges it brings. In each of our examples there is the theme of automated event detection in a signal stream, be it image or acoustic. It is clear that in well-characterized problems, event detectors are readily automated. However, many of the challenges we see in environmental monitoring require iteration and human interaction to be adequately flexible.

The creation of suitable classification algorithms to discriminate the unique vocal signatures of not only species, but individuals in a species is highly desirable. Coupling automated event detection and localization with classifiers that can run on-line and in the field will undoubtedly assist bioacoustics research and enable new research questions to be posed in the future.

In plant phenology, distinguishing the leaves of different species using an imager in a stand of mixed trees has so far not been possible. Additionally, small annual and perennial wildflowers are particularly difficult to detect using color alone and thus other, more sophisticated image processing tools, such as template matching or SIFT, may hold promise.

It has been observed that different techniques work better at different stages of the avian cycle. Therefore, enabling a level of adaptation where the image processing tools could autonomously infer the stage of the cycle and then adapt their processing would be highly desirable. Having the scientist train the system to detect change points, after which the system could respond to its own context classification, is a promising approach.

Whilst automated systems to perform context-aware processing of data from biological sensors are the desirable end-goal, the scientist is likely to be in the loop for the foreseeable future. Thus, interactive environmental sensing will still be required and will necessitate concurrent technological advances in data browsing, data processing, and visualization tools.

As automation of the current challenges we see in interactive environmental sensing become more feasible, challenges will occur in accuracy improvement, self-adaptation, context-aware processing and self-configuration.

6. ACKNOWLEDGEMENTS

We would like to acknowledge Andreas Ali, Travis Collier, Mohammad Rahimi and John Hicks for their contributions to the work described in this paper.

7. REFERENCES

[1] J.Porter, P. Arzberger, HW. Braun, P. Bryant, and S. Gage et al, "Wireless sensor networks for ecology," *BioScience*, vol. 55, no. 7, pp. 561–572, July 2005.

[2] L. Girod, M. Lukac, V. Trifa, and D. Estrin, "The design and implementation of a self-calibrating acoustic sensing platform," in *SenSys '06: Proceedings of the 4th international conference on Embedded networked sensor systems*, 2006.

[3] A. Ali, T. Collier, L. Girod, K. Yao, C.E. Taylor, and D. Blumstein, "An empirical study of collaborative acoustic source localization," in *IPSN '07: Proceedings of The International Conference on Information Processing in Sensor Networks*, 2007.

[4] M.D. Schwartz, R. Ahas, and A. Aasa, "Onset of spring starting earlier across the northern hemisphere," *Global Change Biology*, vol. 12, no. 2, pp. 343–351, 2006.

[5] C. Parmesan, "Ecological and evolutionary responses to recent climate change," *Annual Review of Ecology, Evolution, and Systematics*, vol. 37, no. 1, pp. 637–669, 2006.

[6] E.E. Cleland, I. Chuine, A. Menzel, H.A. Mooney, and M.D. Schwartz, "Shifting plant phenology in response to global change," *Trends in Ecology and Evolution*, vol. 22, no. 7, pp. 357–365, 2007.

[7] C Harris and M Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.

[8] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.