FALL 2020 – COS 397 COMPUTER SCIENCE CAPSTONE PROJECT PROPOSALS

Project Title:

A graphical user tool (BirdSpotter) for integrating machine learning and field survey data for rapid population estimation of colonial nesting birds

Proposed By:

Dr. Cynthia S. Loftin, Wildlife, Fisheries, and Conservation Biology (WFCB)

Team members:

Dr. Roy Turner, Dr. Salimeh Sekeh, Dr. Kate Beard, and Alex Revello (graduate student), School of Computing and Information Science

Logan Kline and Meredith Lewis (graduate students), Lauren Maher (undergraduate) WFCB

Dr. Daniel Hayes, Tony Guay, David Sandilands, School of Forest Resources and Wheatland Geospatial Lab

Brief Description:

Our team's research is investigating the use of remote sensing technologies for collecting population and habitat use information for colonial nesting bird species in coastal environments, while increasing the capacity for collaborative data collection, management, and analysis with agency biologists and citizen scientists. We are targeting seabirds and wading birds that nest in island colonies and coastal rookies in the Gulf of Maine region for this research. State and federal natural resource managers are responsible for annually estimating populations of these species and monitoring their distributions, and to date they have relied on manually interpreting plane-based imagery and walking ground surveys to collect this information. These approaches are limited by resolution of the aerial imagery and the disturbance caused during ground surveys, as well as the expense, inaccessibility of many of the colony locations, and the significant time required to conduct the surveys and interpret the imagery. Our research ultimately will enable wildlife managers to more accurately and rapidly meet their migratory bird management responsibilities and increase their understanding about how emerging computer science technologies can improve their work efficiency, while providing students real-world experiences working with practitioners to develop and apply these technologies in resource management.

The aim of our research is multifaceted. We are 1) identifying strategies that increase colonial nesting bird survey accuracy and efficiency in the region, while reducing data acquisition costs and disturbance to the focal populations during the nesting period; 2) engaging biologists who monitor Maine's coastal wildlife to help us develop and apply of state-of-the-art wildlife survey and data analysis technologies; 3) facilitating transfer of information to biologists by creating workflows, data standards, visualization practices, and tools they can use to incorporate this research into their decision-making; and, 4) sharing the science and technologies with resource managers and citizen scientists via workshops and demonstrations. Our research valuates data collection platforms (plane, UAV, ground surveys), data types (imagery, real-time observations), and data collection protocols (e.g., UAV speed and altitude) for surveying nests and attending birds. We are developing machine learning techniques to estimate numbers of birds while also assessing benefits, risks, and efficacy of the automated technologies, with respect to disturbance, response of the focal species, detectability, and error estimation. Our study is providing rich opportunities for students to engage Maine's biologists and coastal communities in emerging computer science technologies while learning about coastal ecosystems and the science that informs resource management.

We have collected a variety of datasets for estimating colonial nesting bird population sizes, productivity, location-specific nest distributions, and species interactions. During spring 2020, we collaborated with students from **COS 465/565 Data Visualization** to develop data summary and visualization tools to rapidly display ground survey data in maps and graphic summaries. Our project's student team members now are collaborating to develop the machine learning algorithms to automate interpretation of imagery (plane- and UAV-based) captured of nesting birds, and they are also manually interpreting this imagery to validate the algorithms. We are developing population estimates, error assessments, and cost valuations of the various survey and interpretation methods, as well as combining visual summaries developed by the DataVis students and team member Logan Kline in **SIE 510** to rapidly display manually counted birds. *We seek assistance from a team of COS397 students to develop a graphical user tool (BirdSpotter) that obtains and processes data by interacting with other programs to bring these processes together, enabling several functions: 1) spatially display user-selected datasets, including ground surveys, manually interpreted plane- and UAV-captured imagery, and machine-learning output datasets; 2) facilitates real-time application of machine-learning code to identify birds in plane- and UAV-collected images; and, 3) allows a user to select datasets to plot in tables, summary charts, and maps.*

Use of emerging technologies to improve efficiency of population estimation requires that the technologies be easily invoked by biologists and managers. This is a critical need in wildlife conservation, yet to date this need has been poorly addressed. The proposed user interface tool will significantly enhance the ability of our state and federal agency partners to rapidly process and interpret datasets such as ours, to assess the status of the bird populations they manage in the Gulf of Maine. Additionally, the user interface will provide an adaptable tool that may be applied for rapid machine learning interpretation and data summary for other image-based wildlife population assessments.

Goals for the project:

We are interested in a program that allows the user, via a graphical user interface, to interact with a variety of other programs to retrieve, visualize, and process scientific data, in particular, colonial nesting bird imagery. The program would have capabilities that include:

1) Applying user-specified machine learning algorithms to user-selected imagery;
2) Labeling interpretations with species, behavior (e.g., nesting, roosting, flying), and probability of correct assignment;
3) Reporting processing metrics (e.g., speed, error, reliability, repeatability);
4) Summarizing interpreted datasets (e.g., total numbers of birds, colonies, species and behaviors, error) in tabular and graphical summaries (e.g., bar chart, line graph, maps with graduated symbols); and,
5) Providing a framework for rapid comparison among data types, locations, and years to examine spatial relationships and trends in focal species and interpretation reliability.

We have latitude and longitude coordinates for bird observations (species, behaviors; ground surveys, manual and machine-learning interpreted) for multiple years and mosaicked imagery from plane- and UAV-based collection platforms. Team members Logan Kline and Alex Revello are developing the machine-learning code in class and thesis project work. Data are in Excel spreadsheets, attribute tables in GIS shapefiles (that can be exported to .txt files), and raster images (for background displays or for analysis) are .tif files. Faculty and staff team members are providing project oversight, and we are meeting regularly with USFWS and Maine Department of Inland Fisheries and Wildlife biologists, who provide technical assistance with biological information. These collaborators ultimately will use the research products and the proposed interface, hence their engagement is critical to the project's success.

Total Duration/Elapsed Time (24 weeks/two semesters):
Our project is a multi-year project, however, we have data sets that are available for use for developing the user interface during the Capstone project. The duration or elapsed time would be determined by the student team and time they can devote to the work. If selected to partner with a Capstone Project team, our research team would be committed to working with the student team as needed as well as regularly in biweekly project meetings. We also would involve Capstone students in our meetings with professional biologists and in opportunities to present their work in venues with external partners as appropriate.

External Schedules/Deadlines: The anticipated end date for our ongoing project is late 2022. A proposed list of products for which we would develop a schedule with the Capstone student team includes:

1) Develop user friendly interface tool that can be efficiently used to upload, interpret with the collaborator-provided machine learning algorithms, and summarize the interpreted imagery content;
2) Identify appropriate display approaches for the focal datasets and analysis summaries;
3) Understand the limitations and strengths of different types of user interfaces and components and effects on usability and interpretation by different end-users; and,
4) Apply and evaluate efficacy of the final product in a poster/presentation and co-authored manuscript (with the team) for publication.

Expected Project Experiences:

There is opportunity for the students involved in this project in every aspect of this list:

<u>Problem definition</u>
<u>Project scope definition</u>
<u>Design and implementation of research methodology (user interface)</u>
<u>Use of applied statistics</u>
<u>Data analysis</u>
<u>Workflow analysis</u>
<u>Development of functional specifications</u>
<u>Identification of and negotiation for needed project resources</u>
<u>Examination of an unfamiliar technical area</u>
<u>Identification of others' technical expertise</u>
<u>Identification and evaluation of alternatives</u>
<u>Development and presentation of recommendations</u>
<u>Responsibility and accountability for a discrete product</u>
<u>Role definition in a task group and participation in group dynamics</u>

Observation of supervisory activities (e.g., personnel assignment, training, development of procedural guidelines)
Observation of management styles
Observation of organizational politics
Preparation of a manuscript for publication

Recommended Experience:

No particular expertise beyond that of typical senior-level computer science students is anticipated. With respect to programming languages to use, we would expect that Python would be a good choice due to its widespread use, its rich set of libraries for data science (including those we are using for machine learning), its portability across operating systems, and its likely maintainability over time. This also is the language used by the COS 465/565 students in the Data Visualization project. A familiarity with ArcGIS is a plus, but it is not required. If a Web-based frontend is decided upon (with advice from the student project team during initial requirements analysis for the project), then JavaScript programming would likely also be needed.

Expected Outputs/Products and likely requirements:

A desired product is a graphical user interface tool that can be engaged by wildlife biologists who have only rudimentary computer skills (i.e., no programming skills) to interact with a variety of other programs to retrieve, visualize, and process scientific data, including imagery. We anticipate this will require integrating the Python-based data visualization code and machine learning algorithms. While a standalone Windows program would likely meet our needs for the near future, given the anticipated need for widespread use of the data (e.g., ours and similar data collected by others) and programs, a Web-based interface would be desirable. We would decide this during the requirements analysis portion of the project as we meet with and get advice from the student project team.

We do *not* anticipate the need at this stage of our project to develop a server to support multiple simultaneous clients via the Web. However, a Web-based user interface might best be implemented as communicating with a backend system via the network or local sockets; again, we will accept advice about implementation issues such as this from the student project team.

For the backend, as noted, we would expect Python to be a good choice for implementation language and, given the potential users of the system, Windows would be the target operating system. However, given that Python is fairly cleanly-portable across operating systems, being an interpreted language, we do not foresee the target operating system to be an important factor. The same is true for the front-end user interface if a Web-based solution is not selected.

Past Experiences by the Client:

This would be a new system. Other counting software exists (e.g., CountEm); however, some of these are written in technical jargon and can be difficult for end-users to interpret. Writing instructions for ecologists and end-users rather than a technical audience will be an important aspect of this user interface tool.

Proposed Testing Plan:

We do not have particular testing strategies in mind at this point. However, we would expect the team to plan for and include usability tests involving the biologists and others in the extended project team. We can provide test datasets, programs to interface to, and so forth, as well as a computer on which to test the program. We can also facilitate communication between the student team and our professional biologist collaborators.

We do have access to historical imagery that could be used to assess feasibility of using algorithms that were developed on 2019 and 2020 images. Additionally, count data and imagery are available for 2019 and will be available for 2020 by mid-fall semester.

Benefits to UMaine:

Our project is an ongoing example of team science, providing students with an opportunity to work on a real-world problem with resource management professionals to address a critical need in conservation biology with emerging technologies. Our project engages students from diverse disciplines (Ecology and Environmental Sciences, Wildlife, Fisheries and Conservation Biology, Computer Science), at different education levels (undergraduate, MS and PhD graduate students), on two campuses (UMaine Orono and Machias), involves faculty with a variety of experiences and skills sets (wildlife management and conservation biology, spatial ecology, computer science, machine learning, remote sensing and image interpretation), and is a partnership with biologists in state (Maine Department of Inland Fisheries and Wildlife, Maine Department of Conservation) and federal (US Fish and Wildlife Service, US Geological Survey) agencies and a private industry-supported lab (Wheatland Geospatial Lab). Our project has been recognized by the UMS Board of Trustees as an example of student learning in a team science framework. Additionally, our team is participating, by invitation, in a Community of Practice (COP) in Remote Sensing/Machine Learning for Wildlife Survey, a collaboration that engages researchers and practitioners from across North America to work together to share knowledge and experience in deploying remote sensing tools for improving wildlife surveys. This proposed project is unique among those in the COP, as well as will significantly advance the efficiency of our agency partners, while strengthening the connections of UMaine with the public community. Finally, the proposed user interface tool (BirdSpotter) has the potential to be transformative: there are teams elsewhere developing machine-learning algorithms and comparing them with manually interpreted imagery for wildlife surveys, however, there are few examples of user interface tools that enable practitioners to easily and rapidly deploy the algorithms to interpret the imagery and develop population estimates in near real-time. Development of this user interface tool to provide this capability will be ground-breaking for wildlife conservation and management.

Project Sponsors:

Our project is funded by grants and in-kind support from UMaine (RRF and AI Initiative grants; WFCB TAs), US Geological Survey, US Fish and Wildlife Service (USFWS), the Maine Department of Inland Fisheries and Wildlife, and the Maine Department of Conservation. Our project also receives support from the US Geological Survey-Maine Cooperative Fish and Wildlife Research Unit through a Cooperative Agreement with the University of Maine.

Other Resource People:

Our external partners (available to provide feedback on the user interface design and usability) include:

- USFWS (Refuges: Linda Welch, Sara Williams, Rick Schoffler, Brian Benedict; Office of Migratory Bird Management: Mark Koneff)
- Maine Department of Inland Fisheries and Wildlife (Dr. Brad Allen, Danielle D’Auria, Kelsey Sullivan)

- Dr. Glenn Mittlehauser, Maine Bird Observatory

Software/server access required:
Students would need access to a coding application (e.g., Jupyter Notebook, Visual Studio). The project team is currently exploring the storage of datasets on the Advanced Computing Group's servers, which would make access easier by others, including the student project team. The machine learning group will be making use of the newly-acquired, MRI-funded high-performance GPU clusters, also managed by the ACG. We foresee no difficulties providing access to these within reason for the student project members.