My Desktop
Prepare & Submit Proposals
Prepare Proposals in FastLane
New! Prepare Proposals (Limited proposal types)
Proposal Status
Awards & Reporting
Notifications & Requests
Project Reports
Submit Images/Videos
Award Functions
Manage Financials
Program Income Reporting
Grantee Cash Management Section Contacts
Administration
Lookup NSF ID

Preview of Award 1740595 - Annual Project Report

Cover |
Accomplishments |
Products |
Participants/Organizations |
Impacts |
Changes/Problems

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1740595
Project Title:	Collaborative Proposal: EarthCube Integration: ICEBERG: Imagery Cyberinfrastructure and Extensible Building-Blocks to Enhance Research in the Geosciences
PD/PI Name:	Heather Lynch, Principal Investigator
Recipient Organization:	SUNY at Stony Brook
Project/Grant Period:	10/01/2017 - 09/30/2020
Reporting Period:	10/01/2017 - 09/30/2018
Submitting Official (if other than PD\PI):	N/A
Submission Date:	N/A
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	N/A

Accomplishments

* What are the major goals of the project?

Polar geosciences stands at the precipice of a revolution, one enabled by the confluence of cutting edge analytical tools, petabytes of high-resolution imagery, and an ever growing array of high performance computing resources. With these tools at hand, we can look beyond incremental improvements in our understanding of the polar regions; instead, fully capitalizing on these opportunities will radically change our capacity to address priority questions for NSF Geosciences. Near-real time datasets of geological and biological importance at the continental scale are within our reach if we create those critical cyberinfrastructure (CI) components that allow the geosciences community to exploit existing assets and establish a common workflow for reproducible imagery-enabled science. The research objective of this proposal is to understand the biological, geological, and hydrological functioning of the polar regions at spatial scales heretofore beyond the reach of individual PIs, and to develop tools for imagery-enabled science that can be applied globally. The resulting cyberinfrastructure, which we call ICEBERG — Imagery Cyberinfrastructure and Extensible Building-Blocks to Enhance Research in the Geosciences, is an extensible system for coupling open-source image analysis tools with the use of high performance and distributed computing (HPDC) for imagery-enabled geoscience research. Moving from megabytes to petabytes of available imagery requires new computational approaches to interpretation coupled with efficient use of HPDC resources. As the spatial resolution of satellite imagery has continued to shrink, we've seen the convergence of traditional pixel-based remote sensing and computer vision approaches based on object characteristics such as context and texture. Additionally, while the sheer volume of imagery poses challenges for processing, we now have enough training data to develop deep learning algorithms for imagery interpretation that can accelerate the speed of classification and geoscience research. To address these challenges, we propose an Integration project to (1) develop open source image classification tools tailored to high-resolution satellite imagery of the Arctic and Antarctic to be used on HPDC resources, (2) create easy-to-use interfaces to facilitate the development and testing of algorithms for application specific geoscience requirements, (3) apply these tools through four use cases that span the biological, hydrological, and geoscience needs of the polar community, (4) transfer these tools to the larger (non-polar) EarthCube community for continued community-driven development.

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities:

One of the major activities of this year of the project was to arrive at a mutual understanding of the four use cases and their cyberinfrastrcture needs. To do this, we has an all-hands Skype call every 2 weeks. This allowed each domain science group to understand the challenges of the other groups, and for the cyberinfrastructure team to understand what components could be shared. More recently, we established an additional bi-weekly cyberinfrastructure Skype call in which the cyberinfrastucture team (primarily Stony Brook and Rutgers) would discuss some of the more detailed software development and HPC challenges being addressed. We have attached two notebooks describing our experiments to scale up the seal use case and the image geo-location (ASIFT) use case, which represent some of the activities to-date on cyberinfrastructure.

Below we describe the activities of the last year specific to each of the four use cases.

<u>Use Case: Seal Detection (Stony Brook)</u>

- We spent a considerable amount of time early in Year 1 creating a labeled set of imagery for training the convolutional neural network (CNN) to detect and count seals in high-resolution satellite imagery, and designing training and validation sets to test the performance on the CNN.
- We completed several detailed experiments with classification and regression CNNs to determine which of these two approaches may be more accurate for censusing seals.
- We wrote up documentation of our code in jupyter notebooks for transparency and reproducibility.
- We have ported our code to several HPC resources (initially, Stony Brook University's SeaWulf cluster and more recently to XSEDE's Bridges).
- We have been working to optimize the resource allocation for large-scale processing of imagery.

<u>Use Case: Land Cover Classification (NAU)</u>

- We have started working on methods to automatically identify, characterize, and remove atmospheric contributions and shadows from scenes, which would allow for future automation of identifying land cover types. Mark Salvatore visited Heather Lynch on July 12 at Stony Brook University and they walked through the current workflow for atmospheric correction of imagery.
- Brad Spitzbart has started working on an approach to use Spectral Angle Mapper and automated clustering of spectrally similar areas to identify regions that might be used to build a atmospheric correction model.
- In this last year, a Ph.D. student (Helen Eifert) was recruited to work on this project. She started her PhD in August 2018 towards the end of the project year.

<u>Use Case: Image Geo-Location (UC Boulder)</u>

- We have been developing an automated geo-location algorithm to produce high-resolution observational 50-80 year timescale ice changes over Antarctica and Patagonia using collections of historical airborne imagery and modern high-resolution WorldView imagery. In this last year, we have been developing an algorithm based on computer vision and machine-learning techniques to automate the geo-location of un-located or poorly-located historical imagery against high-resolution satellite imagery by matching time-invariant topographic features.
- We have been developing an automated geo-location algorithm to produce high-resolution observational 50-80 year timescale ice changes over Antarctica and Patagonia using collections of historical airborne imagery and modern high-resolution WorldView imagery. In this last year, we have been developing an algorithm based on computer vision and machine-learning techniques to automate the geo-location of un-located or poorly-located historical imagery against high-resolution satellite imagery by matching time-invariant topographic features.
- In this last year, a postdoc (Michael McFerrin) was recruited to work on this project.

<u>Use Case: Hydrology (UCSB)</u>

- We have discussed the existing pipeline for hydrological feature detection and have started working on automating it using existing Python-based tools.In this last year, a Ph.D. student was recruited to work on this project, who is scheduled to start in Fall 2018.

Specific Objectives:

<u>Use Case: Seal Detection</u>

- We have developed a complete HPC pipeline to detect pack-ice seals in high-resolution satellite imagery.
- We have uploaded with this annual report the code and documentation behind the seal detection pipeline. Note that this code has already been used to support another project outside of the four use cases developed through our funding, which speaks to the extensibility of the code for other research groups.
- We have integrate the seal detection pipeline with HPC resources through the ICEBERG middleware we have developed.

<u>Use Case: Image Geo-Location</u>

- We have download and modified an implementation of the ASIFT (Affine- and Scale-Invariant Feature Transform) algorithm to detect shared keypoints between images.
- We have performed initial image-scaling tests to inform the use of the ASIFT algorithm on large satellite imagery and develop tiling strategies for implementation (see uploaded supporting documents).
- We have performed formal strong- and weak-scaling tests to assess algorithm performance and viability on HPC systems (see uploaded supporting documents).

Significant Results:

Use Case: Seal Detection

- We have assembled a 78,000 image training set to train CNNs for seal detection.
- Our best performing architecture achieves over 90% precision and recall when detecting crabeater seal haul outs on a validation set with 11 classes.
- We now have a complete seal detection pipeline running on Bridges that ingests high-resolution satellite imagery and produces shapefiles with patch classification and seal counts.

Use Case: Image Geo-Location

- We have compiled and implemented an initial version of the ASIFT algorithm successfully on satellite and airborne TIF images (see ASIFT Supplementary Materials; Figure 1), with a mix of correct and false-positive matches:
- We have performed initial image-scaling tests on ASIFT executable. Scaled by input image size, the algorithm performs better-than-linearly with respect to memory usage and quadratically with respect to processing time on a multicore desktop

workstation (see ASIFT Supplementary Materials; Figure 2). The primary constraint of HPC implementation will be memory limits, requiring tiling of both source and target images with sizes ranging from 100 megapixels to 2 gigapixels apiece.

- Our next step on this component of the project is to tile the images, as well as efficiently filtering out false-positive keypoint matches using machine-learning algorithms such as RANSAC.

Key outcomes or Other achievements:

*** What opportunities for training and professional development has the project provided?**

In the first year, this project has provided research opportunities for one Ph.D. student at Stony Brook University (Bento Goncalves) and one post-doctoral researcher at the University of Colorado Boulder (Michael MacFerrin). In addition, co-PI Salvatore (NAU) has recruited one graduate student to start in Fall 2018 (Helen Eifert, St. Lawrence University).

Bento Goncalves (Stony Brook) was selected to present his research at an Organized Oral Session at the Polar 2018 conference in Davos, Switzerland, where he met and discussed with other researchers using high resolution satellite imagery to track wildlife in polar regions. He also presented his work as part of Stony Brook Computer Science Department's Computer Vision seminar, where he received valuable feedback and new ideas from the computer vision community. Bento Goncalves was also selected to participate in a Bayesian modelling summer course (May 2018) at the National Socio-Environmental Synthesis Center (SESYNC) in Annapolis, MD, where he improved his Bayesian modelling skills as will be required for the next step of his seal detection pipeline under ICEBERG. Bento has learned a number of technical skills over the last year, specifically in the area of computer vision. Through ICEBERG and the interaction with the cyberinfrastructure team, Bento has also learned a lot about Python software development, documentation, code testing, and reproducible research methods (Jupyter notebooks, etc.).

Co-PI Shantenu Jha (Rutgers) and one of the Ph.D. students in his lab working on ICEBERG (Ioannis Paraskevakos) attended the EarthCube Research Coordination Network Workshop - Advancing the Analysis of High Resolution Topography (A2 HRT) in August 2018. This workshop not only provided networking and professional development for Paraskevakos but it also provided some technical background related to the digital elevation model and geo-location use case of ICEBERG.

*** How have the results been disseminated to communities of interest?**

PI Lynch (Stony Brook) attended the EarthCube AHM in Alexandria, VA and presented a poster at that meeting describing ICEBERG and its progress to date. This was a valuable opportunity to meet with other EarthCube funded projects (e.g., PanGEO).

The Polar 2018 Conference in Davos, Switzerland (June 15-23, 2018) was a significant opportunity to introduce ICEBERG to the polar science community, and to disseminate early results from the seal detection pipeline. Lynch presented ICEBERG to the Scientific Committee for Antarctic Research (SCAR) Standing Committee on Antarctic Data Management (SCADM), where she received valuable feedback and ideas from the assembled data managers. Lynch also presented the details of the seal detection pipeline to the SCAR Action Group on Remote Sensing of Wildlife at Polar 2018, as well as to the Southern Ocean Observing System (SOOS) Censusing Animal Populations from Space (CAPS) working group. Bento Goncalves presented his early results on the seal detection pipeline in an oral session organized by Lynch (and Michelle LaRue, University of Minnesota) focused on the use of high resolution imagery for wildlife detection.

Post-Doc Mike MacFerrin (UC Boulder) participated in an University of Colorado EarthLab Spatial Data Tools workshop in July 2018, and will present initial results of the Image Geo-Location use case at the American Geophysical Union 2018 Fall Meeting.

*** What do you plan to do during the next reporting period to accomplish the goals?**

Having spent a portion of Year 1 recruiting all the personnel who will be involved with ICEBERG, we are now fully staffed and expect to make considerable progress on our project's objectives in Year 2. We will likely complete work on the seal detection use case in Q2 of Year 2 and will therefore be focused on writing up those results for publication. We expect to make considerable progress on the remaining three use cases and hope to complete the initial pan-Antarctic land cover classification by the end of Year 2. We anticipate holding a budgeted ICEBERG science workshop at Stony Brook University towards the end of Year 2, where we will invite all of those directly incolved with ICEBERG as well as representatives from other research groups interested in learning about, and ultimately adopting, the cyberinfrastructure developed by the ICEBERG team.

Supporting Files

Filename	Description	Uploaded By	Uploaded On
ASIFT Supplementary Materials.pdf	Figures relating to the ASIFT algorithm as referenced in the annual report.	Heather Lynch	09/10/2018
goncalves_IACS_JR_2018.pdf	Presentation with more details of the seal use case	Heather Lynch	09/10/2018
Seal Use Case Scaling.pdf	Information about the scaling of the seal use case	Heather Lynch	09/10/2018
ASIFT Use Case Scaling.pdf	Information about the scaling of the ASIFT use case	Heather Lynch	09/10/2018

Products

Books

Book Chapters

Inventions

Journals or Juried Conference Papers
View all journal publications currently available in the NSF Public Access Repository for this award.

The results in the NSF Public Access Repository will include a comprehensive listing of all journal publications recorded to date that are associated with this award.

Licenses

Other Conference Presentations / Papers
B. Goncalves, H.J. Lynch (2018). *Monitoring pack-ice seals from space with deep learning*. POLAR 2018. Davos, Switzerland. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Other Products

Other Publications

Patents

Technologies or Techniques

Thesis/Dissertations

Websites

Participants/Organizations

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked

Name	Most Senior Project Role	Nearest Person Month Worked
Lynch, Heather	PD/PI	6
Spitzbart, Bradley	Other Professional	8
Goncalves, Bento	Graduate Student (research assistant)	12

Full details of individuals who have worked on the project:

Heather Lynch
Email: heather.lynch@stonybrook.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 6

Contribution to the Project: Lead PI responsible for overall project organization and coordination. Supervisor of Bento Goncalves and Brad Spitzbart at Stony Brook, and oversaw the seal use case component of the project.

Funding Support: NSF OPP & GSS CAREER Award No. 1255058 NASA NNX14AC32G

International Collaboration: No
International Travel: Yes, Switzerland - 0 years, 0 months, 10 days

Bradley Spitzbart
Email: bradley.spitzbart@stonybrook.edu
Most Senior Project Role: Other Professional
Nearest Person Month Worked: 8

Contribution to the Project: Worked closely with the Rutgers team to develop code for the four uses cases. Serves as the primary liaison between the domain scientists and the computer scientists on the project.

Funding Support: None other than this award.

International Collaboration: No
International Travel: No

Bento Goncalves
Email: bento.goncalves@stonybrook.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 12

Contribution to the Project: Led the code development of the seal use case.

Funding Support: Additional support provided by the Stony Brook Institute for Advanced Computational Science graduate award

International Collaboration: No
International Travel: Yes, Switzerland - 0 years, 0 months, 10 days

What other organizations have been involved as partners?

Name	Type of Partner Organization	Location
Polar Geospatial Center	Academic Institution	Minneapolis, MN

Full details of organizations that have been involved as partners:

Polar Geospatial Center

Organization Type: Academic Institution
Organization Location: Minneapolis, MN

Partner's Contribution to the Project:
In-Kind Support
Collaborative Research

More Detail on Partner and Contribution: We have worked closely with the Polar Geospatial Center to plan for the movement of imagery from PGC's archive to HPC resources. The PGC has also collated a set of cloud-free imagery to be analyzed for the land cover classification use case.

What other collaborators or contacts have been involved?
Nothing to report

Impacts

What is the impact on the development of the principal discipline(s) of the project?

Each of the four use cases under development are likely to have significant impacts in their respective areas. The use of deep learning to annotate satellite imagery is a growing area and one that is likely to have profound consequences for both biology and geology. For example, our work on the seal use case will provide a detailed case study in the use of deep learning for biological survey, and it will identify some of the cyberinfrastrcture requirements for large-scale deployment. At the project's completion, we will have a robust estimate of the total abundance of pack-ice seals in the Southern Ocean, which is critical for understanding the consumption of krill by pack-ice seals. The land cover classification project will provide detailed land cover classification maps for all the bare rock areas in Antarctica, which will significantly expand our understanding of the continent's geological history. The geo-location use case will allow us to more fully utilize historic airphotos for change detection going back before the satellite era. All of these advances will contribute significant understanding within their respective disciplines.

What is the impact on other disciplines?

There are two types of cross-disciplinary impacts this project will have. One the one hand, the products to be produced will benefit other communities. One example of cross disciplinary impacts is the application of a land cover classification product to biological feature detection. One of the largest sources of false positives in the detection of penguin colonies in high resolution imagery are the appearence of iron-rich rock types that spectrally mimic penguin guano. A detailed high resolution land cover product would improve the ability of penguin biologists to eliminate such false positives. The second type of cross disciplinary impact comes about via the cyberinfrastrcture to be developed, which will provide benefits for all imagery-enabled science questions.

What is the impact on the development of human resources?
Nothing to report.

What is the impact on physical resources that form infrastructure?
Nothing to report.

What is the impact on institutional resources that form infrastructure?
Nothing to report.

What is the impact on information resources that form infrastructure?
Nothing to report.

What is the impact on technology transfer?

Nothing to report.

What is the impact on society beyond science and technology?

Satellite imagery of our planet is captivating to the general public, and this is even more true in the most remote corners of the planet unlikely to ever be visited. The widespread interest in the Arctic and Antarctic digital elevation models (REMA and ArcticDEM) provide one such example of how such large-scale products can drive the public conversation. The products produced by ICEBERG are likely to have similarly large public interest, since they will directly contribute to our ability to address global concerns of environmental change in the polar areas and beyond. The geo-location project, by accurately overlaying historic air photos on modern satellite imagery, will provide a strong visual perspective on long-term environmental change. Such imagery is often a powerful force in terms of the public's understanding of otherwise complex issues.

Changes/Problems

Changes in approach and reason for change
Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them

Our recruitment of a Senior Research Programmer at Rutgers took longer than anticipated but we have now recruited this position. Because other personnel were available to complete the required work in the interim, the impact on the overall progress of the project was minimized.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.