The Journal of Open Source Software

psi-collect: A Python module for post-storm image collection and cataloging

Matthew C. Moretz[1], Daniel Foster[1], John Weber[1], Rinty Chowdhury[1], Shah Nafis Rafique[1], Evan B. Goldstein[2], and Somya D. Mohanty[1]

1 Department of Computer Science, The University of North Carolina at Greensboro **2** Department of Geography, Environment, and Sustainability, The University of North Carolina at Greensboro

DOI: 10.21105/joss.02075

Software

- Review
- Repository
- Archive

Editor: Katy Barnhart

Reviewers:

- @katmratliff
- @conlin-matt
- @chrisleaman

Submitted: 03 February 2020
Published: 01 March 2020

Summary

Major storms along the coastline can damage infrastructure and leave deposits of sediment and debris. Synoptic aerial imagery is often used to assess damage and impacts from storm events. A key source for this imagery in the US is the Emergency Response Imagery (ERI) collected by the the National Geodetic Survey (NGS) Remote Sensing Division of the US National Oceanographic and Atmospheric Administration (National Geodetic Survey, 2020). This imagery aids in recovery efforts as well as rapid assessment of storm impacts along developed and undeveloped coastlines (Madore, Imahori, Kum, White, & Worthem, 2018).

Post-event imagery is typically large, both in terms of the number of individual image files and the size of each file. For example, Hurricane Florence (2018) has over 29,000 JPEG images, with an average size of 7.7 Mb. The first steps for extracting information from these data involve acquiring and processing images. NOAA ERI is currently available as a Web Map Tile Service or via download using a graphical user interface (directing users to the relevant tar and ZIP archive; National Geodetic Survey, 2020). To enable users to download NOAA ERI images via command line for use in reproducible computational workflows, we developed a Python module (`psi-collect`).

The key functionality of `psi-collect` is it allows users to download specific tar and ZIP archives based on storm name, date of image acquisition (day, month, and/or year), and/or image type (JPEG, TIFF). Users can also filter and select specific files using regular expressions. The module includes an automatic resumption feature in the event that a download is interrupted. Each tar and ZIP archive is checked for integrity of contents upon download completion to ensure that data are accurate and intact.

The module also functions as a tool for managing a user's library of images — users can quickly understand which storms they have downloaded. A cataloging tool is also supplied, which allows users to build CSV files that display key information for each image such as image name, acquisition data, file size, and latitude and longitude for each corner of image (extracted from the associated `.geom` file). This catalog can be used for statistical and spatial analyses.

`psi-collect` addresses four issues experienced by researchers working with large collections of NOAA post storm aerial imagery. First, users previously needed to manually navigate to each storm and individually download the multiple image `.tar/ .zip` archives. Download could be done with the graphical user interface, or via another technique (e.g., `wget` on Linux). Even though `wget` is a command line method for retrieving the data, users would still be required to manually navigate and include individual web addresses for each storm archive (e.g., there are 15 individual archives for Hurricane Florence). `psi-collect` dynamically traverses pages and downloads archives when given specific filtering/sub-setting criteria (i.e., downloading all archives from Hurricane Florence).

Second, the individual image archives are large, and downloads often freeze or terminate early. Some browsers or software tools may be capable of resuming partial downloads, but this is case specific and depends on the retrieval workflow. Terminated downloads may require users to determine the missing archive, and re-initiate a download. `psi-collect` implements both automatic resumption and post-download integrity checks to manage download interruption.

Third, we have observed that download speeds of NOAA ERI imagery can be slow. `psi-collect` handles queuing of downloads and cataloging what downloads exist on a user's local computer. These features aid users when multiple archives are needed.

Fourth, `psi-collect` implements a soft locking mechanism allowing users to simultaneously download archives on multiple computers that upload to a single network file system or integrate with a distributed file systems (e.g., Hadoop). The locking system creates a text file with the same named as the archive, but with an additional ending ('xxxx.tar.lock'). This file has information about what user is downloading that archive, download progress, and how large the archive will be when downloading is complete (and the archive is fully uploaded). This file is updated regularly as the download progresses.

Statement of Need

`psi-collect` enables scientists to download NOAA Emergency Response Imagery via Python in a variety of ways (via date, storm) and obtain metrics on downloaded images though the cataloging functions. We envision that `psi-collect` could be used to develop reproducible computational workflows to analyze post-event imagery. For example, images can be used to: assess damage to the built environment (e.g., Thomas, Kareem, & Bowyer, 2014), categorize impact in the context of the Sallenger Jr (2000) Storm Impact Scale (e.g., Liu et al., 2014; Goldstein et al., 2020), evaluate forecasts of storm impact (Morgan, Plant, Srockdon, & Snell, 2019), measure the morphology of storm deposits (e.g., Overbeck, Long, Stockdon, & Birchler, 2015; Lazarus, 2016), and study how human development controls the shape of sediment and debris deposits (e.g., Rogers et al., 2015).

Acknowledgements

We thank Chris Leaman, Katherine Ratliff, Matt Conlin, and Katy Barnhart for their thoughtful reviews and contributions to the code. EBG acknowledges funding from NSF (Award #1939954), The Leverhulme Trust (RPG-2018-282), and an Early-Career Research Fellowship from the Gulf Research Program of the National Academies of Sciences, Engineering, and Medicine. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Gulf Research Program of the National Academies of Sciences, Engineering, and Medicine.

References

Goldstein, E., Lazarus, E., Beuzen, T., Williams, H., Limber, P., Cohn, N., Taylor, L., et al. (2020). *Labels for hurricane florence (2018) emergency response imagery from noaa*. figshare. doi:10.6084/m9.figshare.11604192.v1

Lazarus, E. D. (2016). Scaling laws for coastal overwash morphology. *Geophysical Research Letters*, *43*(23), 12–113. doi:10.1002/2016GL071213

Liu, S. B., Poore, B. S., Snell, R. J., Goodman, A., Plant, N. G., Stockdon, H. F., Morgan, K. L., et al. (2014). USGS iCoast–did the coast change? Designing a crisis crowdsourcing

app to validate coastal change models. In *CSCW companion '14: Proceedings of the companion publication of the 17th acm conference on computer supported cooperative work & social computing* (pp. 17–20). doi:10.1145/2556420.2556790

Madore, B., Imahori, G., Kum, J., White, S., & Worthem, A. (2018). NOAA's use of remote sensing technology and the coastal mapping program. In *OCEANS 2018 mts/ieee charleston* (pp. 1–7). IEEE. doi:10.1109/OCEANS.2018.8604932

Morgan, K. L., Plant, N. G., Srockdon, H., & Snell, R. J. (2019). ICoast–did the coast change?: Storm-impact model verification using citizen scientists. In *Coastal sediments 2019* (pp. 1424–1438). World Scientific. doi:10.1142/9789811204487_0124

National Geodetic Survey, N. (2020). *Emergency response imagery*. Retrieved from https://storms.ngs.noaa.gov

Overbeck, J. R., Long, J. W., Stockdon, H. F., & Birchler, J. J. (2015). Enhancing evaluation of post-storm morphologic response using aerial orthoimagery from hurricane sandy. In *Coastal sediments 2015*. World Scientific. doi:10.1142/9789814689977_0250

Rogers, L. J., Moore, L. J., Goldstein, E. B., Hein, C. J., Lorenzo-Trueba, J., & Ashton, A. D. (2015). Anthropogenic controls on overwash deposition: Evidence and consequences. *Journal of Geophysical Research: Earth Surface*, *120*(12), 2609–2624. doi:10.1002/2015JF003634

Sallenger Jr, A. H. (2000). Storm impact scale for barrier islands. *Journal of Coastal Research*, 890–895. Retrieved from https://www.jstor.org/stable/4300099

Thomas, J., Kareem, A., & Bowyer, K. W. (2014). Automated poststorm damage classification of low-rise building roofing systems using high-resolution aerial imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(7), 3851–3861. doi:10.1109/TGRS.2013.2277092