

A semi-automated workflow for processing historic aerial photography

Nathan Piekielek^{a, *}

^a *The Pennsylvania State University, University Libraries*

* Corresponding author

Keywords: aerial photography, computer vision, big historical geodata

Abstract:

Libraries, museums and archives were the original big geospatial information repositories that to this day house thousands to millions of resources containing research-quality geographic information. However, these print resources (and their digital surrogates), are not easily incorporated into the contemporary research process because they are not structured data that is required of web-mapping and geographic information system tools. Fortunately, contemporary big data tools and methods can help with the large-scale conversion of historic resources into structured datasets for mapping and spatial analysis.

Single frame historic aerial photographs captured originally on film (hereafter “photographs”), are some of the most ubiquitous and information-rich geographic information resources housed in libraries, museums and archives. Photographs authentically encoded information about past places and time-periods without the thematic focus and cartographic generalization of historic print maps. As such, they contain important information in nearly every category of base mapping (i.e. transportation networks, populated places etc.), that is useful to a broad spectrum of research projects and other applications. Photographs are also some of the most frustrating historic resources to use due to their very large map-scale (i.e. small geographic area), lack of reference information and often unknown metadata (i.e. index map, flight altitude, direction etc.).

The capture of aerial photographs in the contiguous United States (U.S.) became common in the 1920s and was formalized in government programs to systematically photograph the nation at regular time intervals beginning in the 1930s. Many of these photography programs continued until the 1990s meaning that there are approximately 70 years of “data” available for the U.S. that is currently underutilized due to inaccessibility and the challenges of converting photographs to structured data. Large collections of photographs include government (e.g. the U.S. Department of Agriculture Aerial Photography Field Office “The Vault” – over 10 million photographs), educational (e.g. the University of California Santa Barbara Library – approximately 2.5 million photographs), and an unknown number non-governmental organizations (e.g. numerous regional planning commissions and watershed conservation groups). Collectively these photography resources constitute an untapped big geospatial data resource.

U.S. government photography programs such as the National Agricultural Imagery Program continued and expanded in the digital age (i.e. post early 2000s), so that not only is there opportunity to extend spatial analyses back in time, but also to create seamless datasets that integrate with current and expected future government aerial photography campaigns. What is more, satellite imagery sensors have improved to the point that there is now overlap between satellite imagery and aerial photography in terms of many of their technical specifications (i.e. spatial resolution etc.). The remote capture of land surface imagery is expanding rapidly and with it are new opportunities to explore long-term land-change analyses that require historical datasets.

Manual methods to process photographs are well-known, but are too labour intensive to apply to entire photography collections. Academic research on methods to increase the discoverability of photographs and convert them to geospatial data at large-scale has to date been limited (although see the work of W. Karel et al.). This presentation details a semi-automated workflow to process historic aerial photographs from U.S. government sources and compares the workflow and results to existing methods and datasets. In a pilot test area of 94 photographs in the U.S. state of Pennsylvania, the workflow was found to be nearly 100-times more efficient than commonly employed alternatives while achieving greater horizontal positional accuracy. Results compared favourably to contemporary digital aerial photography data products, suggesting that they are well-suited for integration with contemporary datasets. Finally, initial results of the workflow were incorporated into several existing online discovery and sharing platforms that will be highlighted in this presentation. Early online usage statistics as well as direct interaction with users demonstrates the broad interest and high-impact of photographs and their derived products (i.e. structured geospatial data).