Supporting Historical Photo Identification with Face Recognition and Crowdsourced Human Expertise (Extended Abstract)*

Vikram Mohanty[1] , **David Thames**[1,2] , **Sneha Mehta**[1] and **Kurt Luther**[1]
[1]Department of Computer Science and Center for HCI, Virginia Tech, USA
[2]Google, USA
{vikrammohanty, davidcthames, sudo777, kluther}@vt.edu

Abstract

Identifying people in historical photographs is important for interpreting material culture, correcting the historical record,and creating economic value, but it is also a complex and challenging task. In this paper, we focus on identifying portraits of soldiers who participated in the American Civil War (1861–65). Millions of these portraits survive, but only 10–20% are identified. We created Photo Sleuth, a web-based platform that combines crowdsourced human expertise and automated face recognition to support Civil War portrait identification. Our mixed-methods evaluation of Photo Sleuth one month after its public launch showed that it helped users successfully identify unknown portraits.

1 Introduction

Identifying people in historical photographs provides significant cultural, historical, and economic value. From a cultural perspective, it can help recognize contributions of marginalized groups, as when the only female African American scientist photographed at a 1971 biology conference was identified [Fortin, 2018b]. Identification can also correct the historical record, as when James Bradley, author of *Flags of Our Fathers*, was convinced by visual evidence that his father was not pictured in the iconic World War II photo of US Marines at Iwo Jima [Schmidt, 2018]. Identification can also create significant economic value, as when a photo of American outlaw Billy the Kid, purchased at flea market for $10, was estimated to be worth millions of dollars [Fortin, 2018a].

However, identifying people in historical photos is complex and challenging, and researchers lack adequate technological support. The current research practices employed by historians, antiques dealers, and collectors for identifying portraits are largely manual and often time-consuming. These practices involve manually scanning through hundreds of low-quality photographs, military records, and reference books, which can often be tedious and frustrating, and lacks any guarantee of success. Automated face recognition algorithms can support this effort, but are not widely used by historical photo experts, and are often insufficient for solving the problem on their own. Many studies have compared face recognition algorithms to a human baseline, with mixed results [Blanton *et al.*, 2016; Best-Rowden *et al.*, 2014; Kemelmacher-Shlizerman *et al.*, 2016]. Further, historical photographs add unique challenges as they are often achromatic, low resolution, and faded or damaged, which might result in loss of useful information for identification.

In this paper, we present Photo Sleuth[1], a web-based platform that combines crowdsourced human expertise and automated face recognition to support historical portrait identification. We introduce a novel person identification pipeline in which users first identify and tag relevant visual clues in an unidentified portrait. The system then suggests filters based on these tags to narrow down search results of identified reference photos. Finally, the user can carefully inspect the narrowed search results, sorted using automatic face recognition, to make a potential identification. This pipeline also bootstraps crowdsourced user contributions to grow the site's database of reference images in a sustainable way, increasing the likelihood of a potential match in the future. Photo Sleuth initially focuses on identifying portraits from the American Civil War (1861–65), the first major conflict to be extensively documented through photographs. An estimated three million soldiers fought in the war and most of them had their photos taken at least once. After 150 years, millions of these portraits survive in museums, libraries, and individual collectors, but the identities of most have been lost.

We publicly launched Photo Sleuth in 2018 and conducted a mixed-methods evaluation of its first month of usage, including content analysis of uploaded photos and expert review of user identifications. We found that the system helped users identify dozens of unknown portraits. Additionally, Photo Sleuth's pipeline encouraged users to voluntarily add hundreds of identified portraits to aid future research, suggesting a sustainable model for long-term participation. Our primary contributions are: (1) a novel person identification pipeline combining crowdsourcing and face recognition; (2) a web-based tool and online community, Photo Sleuth, demonstrating this approach, and (3) a mixed-methods evaluation of

*This paper is an abridged version of a paper titled "Photo Sleuth: Combining Human Expertise and Face Recognition to Identify Historical Portraits" [Mohanty *et al.*, 2019b] that won the Best Paper Award at ACM IUI 2019.

[1]http://www.civilwarphotosleuth.com

Portrait Sleuth after one month of deployment with real users. We also discuss implications for crowd–AI interaction.

2 Related Work

2.1 Person Identification in Photographs

AI-based face recognition algorithms are finding use in many real-world applications. However, these face recognition algorithms often produce many false positives [Press Association, 2018; Singer, 2018], partly due to gender and skin type biases [Raji and Buolamwini, 2019]. Studies have compared face recognition algorithms to human baselines, and some show that human performance is superior [Blanton *et al.*, 2016; Best-Rowden *et al.*, 2014; Kemelmacher-Shlizerman *et al.*, 2016], or suggest that optimal face recognition can be achieved by fusing humans and machines [Phillips *et al.*, 2018]. The workflow of Photo Sleuth prevents face recognition *per se* from making the final identification decision, instead deferring to human judgment.

It is estimated that at least four million Civil War-era portraits survive today, of which 10–20% are already identified [Coddington, 2013]. Civil War portrait identification or "photo sleuthing" typically requires extensive skill and domain expertise, from identifying obscure uniform insignia and weapons [Martinez, 2012] to reviewing thousands of potential facial matches [Luther, 2015]. Photo Sleuth attempts to ease the sleuthing process by bringing together a large repository of soldier portraits, military service records, and visual annotations in a workflow appropriate for both novices and experts.

2.2 Crowdsourced History and Image Analysis

Research on crowdsourcing systems with applications to historical research has largely been limited to transcription projects (e.g., [Williams *et al.*, 2014; Hansen *et al.*, 2013]). While person identification is a more complex task than text transcription and requires more domain knowledge, we draw inspiration from the approaches these projects take to designing interfaces that help crowds visually analyze historical primary sources.

Research involving crowdsourced image analysis often focuses on identifying everyday objects, transcribing text, or other tasks requiring only basic knowledge. Various techniques have been employed for crowdsourcing analysis of unfamiliar visual material in a systematic way, such as combining crowds with computer vision to annotate bus stops and sidewalk accessibility issues in Google Street View images [Hara *et al.*, 2013] or asking volunteer crowds to compare photos of missing and found pets to reunite them with their owners after a disaster [Barrenechea *et al.*, 2015].

Systems like Flock [Cheng and Bernstein, 2015] and Tropel [Patterson *et al.*, 2015] use crowdsourcing to build hybrid crowd–machine learning classifiers. Due to scale and complexity issues, a person identification task cannot be seen as a multi-label or extreme classification problem. Since these approaches require a user to define the prediction task and example labeled data, they cannot be directly applied to a person identification task.

3 System Description

Photo Sleuth is an online platform for identifying unknown people in Civil War-era portraits. It is a Python/Django web application that uses a PostgreSQL database for data storage, Amazon S3 for image storage, and Heroku for cloud hosting. Photo Sleuth allows users to upload photos, tag them with visual clues, and connect them to profiles of Civil War soldiers with detailed military records.

Historical person identification can be seen as *finding a needle in the haystack*. Photo Sleuth's novel pipeline has three components: (1) building the haystack, (2) narrowing down the haystack, and (3) finding the needle.

3.1 Building the Haystack

Reference Database. Photo Sleuth's initial reference database contains over 20,000 identified Civil War soldier portraits from public sources like the US Army Military History Institute and the US Library of Congress, as well as other private sources. This is just a small proportion of the four million photos estimated to exist today. Therefore, a more comprehensive archive with more reference photos and identities would boost Photo Sleuth's purpose of identifying soldiers, and necessitates *building a haystack*.

Photo Upload. A user begins the identification process by uploading a photograph with a mandatory front view and an optional back view. The user is also encouraged to provide the original source of the photo. We use Microsoft Azure's Face API to detect a face in the photograph at the time of uploading. Photo Sleuth does not yet support photos with multiple faces.

Photo Metadata. Next, the user tags metadata related to the photograph, if available, such as the photo format, inscriptions on the front and back view of the photo, and the photographer's name and studio location. This metadata can offer insights into the subject's hometown, military unit, or name, both improving the search filters and providing useful annotations for researchers.

Visual Tags. Our system then gathers information about visual evidence e.g., *Coat Color*, *Chevrons*, *Shoulder Straps*, *Collar Insignia*, or *Hat Insignia*. These visual tags are mapped on to the soldier's military service information, which provides a useful search parameter. More tags improve the relevance of the candidate pool, and thus reduce the number of false positives.

Bootstrapping and Ownership. Photo Sleuth adds the photo along with this information into the reference database, irrespective of identity, while displaying authorship credentials to the user. These photos enrich the database for potentially identifying future uploads. Previous work suggests attribution is an important incentive for crowds conducting original research [Luther *et al.*, 2009].

3.2 Narrowing down the Haystack

Search Filters. A major challenge in person identification tasks is the size of the candidate pool. Larger pools mean greater possibilities for false positives. Photo Sleuth reduces the likelihood of wrong identifications by generating search filters based on the visual evidence tagged by the user. These search filters are based on military service details that would otherwise be unknown to a novice user and therefore, built using domain expertise. The military records used by the filters come from a variety of sources, including the US National Park Service Soldiers and Sailors Database. We scraped the full military service record for every identified soldier portrait in our database, along with, in many cases, vital records and biographical details. This allows for users to filter by visual clues that would only be applicable for a snapshot of a soldier's career.

For example, if the user tagged *Hat Insignia* with a hunting horn, the system would recommend the "Infantry" branch filter, whereas *Shoulder Straps* with two stars would suggest the "Major General" rank filter. These filters narrow down the search pool to all soldiers who might ever have held these positions, including promotions, demotions, and transfers. Our system shows all search filters to the users, allowing expert users to make manual refinements. Photo Sleuth's interface also scaffolds domain knowledge by preventing users from applying search filters that contradict each other.

Facial Similarity. Photo Sleuth augments the above search filters with facial similarity filtering via Microsoft Azure's Face API. Our initial tests with identified Civil War photos showed that this API yields near-perfect recall at a 0.50 similarity confidence threshold; i.e., retrieved search results at this level almost always include the correct results. However, its poor precision means many other similar-looking false positives also show up in the search results.

The search filters create a reduced search space in which face recognition looks for similar-looking photos of the query image. This complementary interaction between military records and facial similarity ensures that the most accurate information is retained in the search space.

3.3 Finding the Needle

Search Results. The search results page displays all the soldier portraits who satisfy the search filters and have a facial similarity score of 0.50 and above with the query photo, sorted by similarity (see Figure 1). The user has the option to hide as-yet unidentified photos. The search results show military record highlights next to the names and photos. The user can then closely investigate the most promising search results before making the final decision of the soldier's identity. The user can also add new names and service records to the database if that soldier profile has not yet been added. In order to prevent misinformation being spread and promote cross-verification, all users are required to follow the entire workflow, even for photos whose identities they believe they already know. In such cases, the user is asked to provide the source of identification.

User Review. Users who find a potential match among the search results can closely inspect the two photos via a "Comparison" interface. The interface provides separate zoom/pan controls and also displays the service records of the reference photo to provide a broader context of who the soldier might be. Notably, the system hides the facial similarity confidence scores for verifying two faces to avoid biasing the user. If

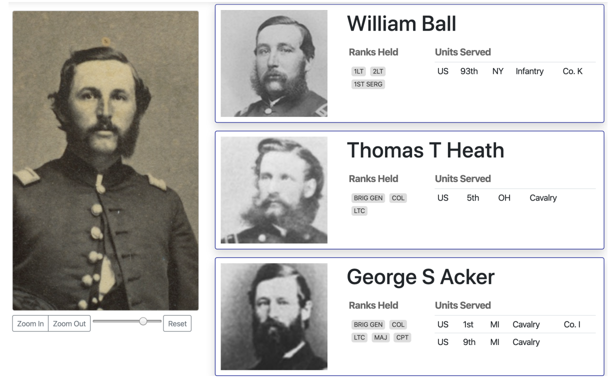

Figure 1: Detail view of Photo Sleuth's search results page, showing the unidentified soldier photo (left) and search results of identified reference images, sorted by facial similarity (right). The top result (William Ball) is the correct match.

the user is confident about the photo being a match, they can click on an "Identify" button to link the query photo to the soldier's profile and receive an "identifier" attribution. The user can also undo these identifications, if desired.

4 Evaluation

We released Photo Sleuth to the public on August 1, 2018. We examined website logs for user-uploaded photos for the first month. During this period, 612 users registered free accounts on the website, and had uploaded 2012 photos.[2]

Log Analysis. The photos were categorized as *identified* and *unidentified*. The identified photos were further categorized into *a) pre-identified photos* (i.e. identities of these photos were known to the user prior to uploading), and *b) post-identified photos* (i.e. photos matched by the user to an existing identified photo using Photo Sleuth's photo matching workflow).

Content Analysis. In order to understand the extent to which face recognition supported a user's identification decision, we checked for the presence of prior *name inscriptions* in the front or back views of the photo. Inscriptions prompt an easy decision on the user's behalf to match the photo with a search result displaying the same name. We also examined whether any of the user-uploaded photos was a *replica* (i.e. an exact duplicate photo) of an identified reference photo of the matched soldier. Based on our findings, we divided the soldier identities with post-identified photos into four subcategories: *a) inscription and replica, b) inscription but no replica, c) replica but no inscription,* and *d) no replica and no inscription*.

Backtracing. For a randomly chosen sample in each of these categories, we backtraced (reconstructed) the identification workflow to re-match a post-identified photo to visualize the user's experience when posed with the search results under the original conditions.

[2]As of March 2020, the site has over 14,000 registered users and over 32,000 photos, of which over 11,000 have been added by users.

Expert Review. An expert Civil War photo historian (and a co-author of this paper) reviewed all post-identified photos added by users and evaluated whether they were correctly identified or not. Based on the expert's response, we consider the matches to be either *positive* matches (Likert-scale ratings of 3–4) or *negative* matches (ratings of 1–2).

In the full version of this paper [Mohanty *et al.*, 2019b], we also report on detailed user experiences with Photo Sleuth, based on interviews with nine active users.

5 Findings

Our logs analysis showed 2012 photos uploaded in the first month, of which 1632 photos were front views and 380 photos were back views. The number of identified photos (560) was similar to the number of unidentified ones (602).

We also observed from our logs that users provided one or more tags for the majority of both identified (445 out of 560) and unidentified (401 out of 602) photos, even though this step is entirely optional. For the tagged photos, users added an average and median of 5 tags per photo. These tags were related to both the photo's metadata (*Photo Format, Photographer Location*, etc.) and the visual evidence in the photos (*Coat Color, Shoulder Straps*, etc.). *Coat Color* and *Shoulder Straps* were the most commonly tagged visual evidence, which the system uses to reduce search results by filtering military records by army side and officer rank, respectively.

We found that users successfully used the system's search workflow to identify unknown photos. Of the 560 identified photos, 441 were pre-identified (i.e., already identified) and 119 were post-identified (i.e., new identifications). These 119 post-identified photos were matched to 88 soldiers with a prior photo in the reference archive.

We observed from backtracing that the matched identity did not always appear as the top search result. Out of 119 post-identified photos, 11 did not have matching identities in the top 50 search results, while 19 matched identities in the top 50 but not the top 1 search result. This suggests users confirmed a match only after carefully analyzing the search results beyond the top few ones.

The expert analyzed all 88 soldiers identities matched with the post-identified photos, and found that all 30 identities with at least one replica were positively matched. Additionally, 20 of the 21 identities with an inscription but no replica were correctly matched. We considered the final category — identities without inscriptions or replicas — to be the most difficult. Out of 37 identities in this category, the expert found 25 identities to be positively matched. Thus, the expert considered the majority of the user-generated identifications in all categories to be positive matches.

6 Discussion

We deliberately decided not to allow the Photo Sleuth system *per se* to directly identify any photos. Although this feature is one of our most persistent user requests, examples from popular media show the danger of a fully automated approach [Press Association, 2018; Singer, 2018]. Instead, the system suggests potential matches largely driven by objective user tagging, and hides quantitative confidence levels. The face recognition algorithm influences results in a more subtle way, by filtering out low-confidence matches and sorting the remainder. This approach emphasizes agency, but at the cost of increased requirements for human attention per image. Because Photo Sleuth helps users quickly identify a much more relevant set of candidates compared to traditional research methods, participants did not seem to view this attention requirement as a major drawback.

This human-led, AI-supported approach to person identification is further emphasized in our design decision to attribute individual users as responsible for particular identifications. This approach aims to promote accountability through social translucence [Erickson and Kellogg, 2000], and to recognize the achievements of conducting original research, as recommended by prior work [Luther *et al.*, 2009]. It also aligns with traditions of expert authentication in the art and antiquarian communities.

Beyond agency and accountability, there are also accuracy-driven reasons to support a human–AI collaborative pipeline. Our evaluation found 30 cases where the automated face recognition failed to rank a correct match as the top search result — or even in the top 50 results — yet users nevertheless found the correct (buried) match. While AI struggles with this "last mile" problem, humans provide complementary strengths in providing close inspections of shortlisted candidates. In follow-up work, we have developed an extension of Photo Sleuth called Second Opinion [Mohanty *et al.*, 2019a] that leverages cognitive science principles to help novice crowds systematically analyze top search results from AI-based face recognition. Our initial results suggest that experts benefit from this crowd–AI hybrid when conducting photo investigations.

In the full paper [Mohanty *et al.*, 2019b], we also discuss building a sustainable model for volunteer contributions, fostering original research while preventing misinformation, and enhancing the accuracy of person identification systems.

7 Conclusion

We address the challenge of identifying people in historical portraits with a novel person identification pipeline that combines crowdsourced human expertise and automated face recognition with contextual information to help users identify unknown Civil War soldier portraits. We demonstrate this approach by building and launching a public web platform, Photo Sleuth, on top of this pipeline. By enabling the identification of dozens of unknown photos, our work opens doors for exploring new ways for building person identification systems that look beyond face recognition alone and leverage the complementary strengths of human and artificial intelligence.

Acknowledgments

We wish to thank Ron Coddington, Paul Quigley, Nam Nguyen, Abby Jetmundsen, Ryan Russell, Natalie Robinson, and our study participants. This research was supported by NSF IIS-1651969 and IIS-1527453 and a Virginia Tech ICTAS Junior Faculty Award.

References

[Barrenechea *et al.*, 2015] Mario Barrenechea, Kenneth M Anderson, Leysia Palen, and Joanne White. Engineering crowdwork for disaster events: The human-centered development of a lost-and-found tasking environment. In *2015 48th Hawaii International Conference on System Sciences*, pages 182–191. IEEE, 2015.

[Best-Rowden *et al.*, 2014] Lacey Best-Rowden, Hu Han, Charles Otto, Brendan F Klare, and Anil K Jain. Unconstrained face recognition: Identifying a person of interest from a media collection. *IEEE Transactions on Information Forensics and Security*, 9(12):2144–2157, 2014.

[Blanton *et al.*, 2016] Austin Blanton, Kristen C Allen, Timothy Miller, Nathan D Kalka, and Anil K Jain. A comparison of human and automated face verification accuracy on unconstrained image sets. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 161–168, 2016.

[Cheng and Bernstein, 2015] Justin Cheng and Michael S Bernstein. Flock: Hybrid crowd-machine learning classifiers. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 600–611. ACM, 2015.

[Coddington, 2013] Ron Coddington. Military Images Magazine https://www.battlefields.org/learn/articles/military-images-magazine, Oct 2013.

[Erickson and Kellogg, 2000] Thomas Erickson and Wendy A. Kellogg. Social translucence: an approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.*, 7(1):59–83, March 2000.

[Fortin, 2018a] Jacey Fortin. A Photo of Billy the Kid Bought for $10 at a Flea Market May Be Worth Millions. *The New York Times*, January 2018.

[Fortin, 2018b] Jacey Fortin. She was the only woman in a photo of 38 scientists, and now she's been identified. *The New York Times*, Mar 2018.

[Hansen *et al.*, 2013] Derek L. Hansen, Patrick J. Schone, Douglas Corey, Matthew Reid, and Jake Gehring. Quality control mechanisms for crowdsourcing: peer review, arbitration & expertise at familysearch indexing. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 649–660, New York, NY, USA, 2013. ACM.

[Hara *et al.*, 2013] Kotaro Hara, Vicki Le, and Jon Froehlich. Combining crowdsourcing and google street view to identify street-level accessibility problems. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 631–640. ACM, 2013.

[Kemelmacher-Shlizerman *et al.*, 2016] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The megaface benchmark: 1 million faces for recognition at scale. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4873–4882, 2016.

[Luther *et al.*, 2009] Kurt Luther, Scott Counts, Kristin B. Stecher, Aaron Hoff, and Paul Johns. Pathfinder: An online collaboration environment for citizen scientists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 239–248, New York, NY, USA, 2009. ACM.

[Luther, 2015] Kurt Luther. Blazing a Path From Confirmation Bias to Airtight Identification. *Military Images*, 33(2):54–55, 2015.

[Martinez, 2012] Ramona Martinez. Unknown No More: Identifying A Civil War Soldier. *NPR*, April 2012.

[Mohanty *et al.*, 2019a] Vikram Mohanty, Kareem Abdol-Hamid, Courtney Ebersohl, and Kurt Luther. Second opinion: Supporting last-mile person identification with crowdsourcing and face recognition. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 7, pages 86–96, 2019.

[Mohanty *et al.*, 2019b] Vikram Mohanty, David Thames, Sneha Mehta, and Kurt Luther. Photo Sleuth: Combining human expertise and face recognition to identify historical portraits. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 547–557. ACM, 2019.

[Patterson *et al.*, 2015] Genevieve Patterson, Grant Van Horn, Serge Belongie, Pietro Perona, and James Hays. Tropel: Crowdsourcing detectors with minimal training. In *Third AAAI Conference on Human Computation and Crowdsourcing*, 2015.

[Phillips *et al.*, 2018] P Jonathon Phillips, Amy N Yates, Ying Hu, Carina A Hahn, Eilidh Noyes, Kelsey Jackson, Jacqueline G Cavazos, Géraldine Jeckeln, Rajeev Ranjan, Swami Sankaranarayanan, et al. Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences*, page 201721355, 2018.

[Press Association, 2018] Press Association. Welsh police wrongly identify thousands as potential criminals. *The Guardian*, May 2018.

[Raji and Buolamwini, 2019] Inioluwa Deborah Raji and Joy Buolamwini. Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. 2019.

[Schmidt, 2018] Michael S. Schmidt. 'Flags of Our Fathers' author now doubts his father was in Iwo Jima photo. *The New York Times*, Jan 2018.

[Singer, 2018] Natasha Singer. Amazon's facial recognition wrongly identifies 28 lawmakers, A.C.L.U. says. *The New York Times*, Jul 2018.

[Williams *et al.*, 2014] Alex C Williams, John F Wallin, Haoyu Yu, Marco Perale, Hyrum D Carroll, Anne-Francoise Lamblin, Lucy Fortson, Dirk Obbink, Chris J Lintott, and James H Brusuelas. A computational pipeline for crowdsourced transcriptions of ancient greek papyrus fragments. In *2014 IEEE International Conference on Big Data (Big Data)*, pages 100–105. IEEE, 2014.