# SleuthTalk: Addressing the Last-Mile Problem in Historical Person Identification with Privacy, Collaboration, and Structured Feedback

Liling Yuan

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science and Applications

Kurt Luther, Chair
Clifford A Shaffer
Christopher L North

May 7, 2021
Blacksburg, Virginia

Keywords: Collaboration, Crowdsourcing, Person Identification, Civil War Photo, Privacy, Structured feedback

# SleuthTalk: Addressing the Last-Mile Problem in Historical Person Identification with Privacy, Collaboration, and Structured Feedback

Liling Yuan

(ABSTRACT)

Identifying people in historical photographs is an important task in many fields, including history, journalism, genealogy, and collecting. A wide variety of different methods, such as manual analysis, facial recognition, and crowdsourcing, have been used to identify the unknown photos. However, because of the large numbers of candidates and the poor quality or lack of source evidence, accurate historical person identification still remains challenging. Researchers especially struggle with the "last mile problem" of historical person identification, where they must make a selection among a small number of highly similar candidates. Collaboration, including both human-AI collaboration and collaboration within human teams, has shown the advantages of improving data accuracy, but there is lack of research about how we can design a collaborative workspace to support the historical person identification. In this work, we present SleuthTalk, a web-based collaboration tool integrated into the public website Civil War Photo Sleuth which addresses the last-mile problem in historical person identification by providing support for shortlisting potential candidates from face recognition results, private collaborative workspaces, and structured feedback interfaces. We evaluated this feature in a mixed-method study involving 6 participants, who spent one week each using SleuthTalk and a comparable social media platform to identify an unknown photo. The results of this study show how our design helps with identifying historical photos in a collaborative way and suggests directions for improvement in future work.

# SleuthTalk: Addressing the Last-Mile Problem in Historical Person Identification with Privacy, Collaboration, and Structured Feedback

Liling Yuan

(GENERAL AUDIENCE ABSTRACT)

Identifying people in historical photographs is an important task in many fields, including history, journalism, genealogy, and collecting. A wide variety of different methods, such as manual analysis, facial recognition, and crowdsourcing, have been used to identify the unknown photos. However, because of the large numbers of candidates and the poor quality or lack of source evidence, accurate historical person identification still remains challenging. Researchers especially struggle with the "last mile problem" of historical person identification, where they must make a selection among a small number of highly similar candidates. Collaboration, including both human-AI collaboration and collaboration within human teams, has shown the advantages of improving data accuracy, but there is lack of research about how we can design a collaborative workspace to support the historical person identification. In this work, we present SleuthTalk, a web-based collaboration tool integrated into the public website Civil War Photo Sleuth which addresses the last-mile problem in historical person identification by providing support for shortlisting potential candidates from face recognition results, private collaborative workspaces, and structured feedback interfaces. We evaluated this feature in a mixed-method study involving 6 participants, who spent one week each using SleuthTalk and a comparable social media platform to identify an unknown photo. The results of this study show how our design helps with identifying historical photos in a collaborative way and suggests directions for improvement in future work.

# Acknowledgments

First and foremost, I'd like to give my appreciation to my advisor, Dr. Kurt Luther. Thank you for the professional and invaluable guidance and continued support. As a student who is new to research, you are patient to teach me how to think and solve problems.

In addition, I would like to thank my committee members, Dr. Cliff Shaffer and Dr. Chris North, for their precious time and excellent suggestions.

I would also like to give my appreciation to Vikram Mohanty for his patience and great help in this project.

And lastly, I'd like to express my sincere gratitude to my parents, Yi Hong and Chi Yuan, for their wholehearted unconditional love and encouragement in my life.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Identifying people in photographs is an important task in many fields, including historical scholarship, investigative journalism, human rights work, and law enforcement. Investigators use a variety of different methods to perform person identification tasks. Researchers have worked on using gait, or height, color, gender to identify the people through surveillance video [17, 20]. Facebook has used facial recognition to tag users in pictures [4]. Law enforcement officials in many jurisdictions employs face recognition to identify potential suspects. However, because of the large numbers of potential candidates and the poor quality or lack of source evidence, person identification has high false positive rates, and many people are misidentified. The entire process of identifying people from photographs remains challenging across domains.

To help with improving the accuracy of person identification, some researchers are turning to crowdsourcing to identify photos and reduce the false positives. This approach has shown promising results. For example, in the domain of history, Willever-Farr and Forte [39] found that two other online historical resources, Ancestry.com and Find-a-Grave, used the crowd's wisdom to collect biographical information in order to maintain high standards of quality. But at the same time, crowdsourced misidentifications can also lead to serious consequences, especially for modern photos. For example, after the U.S. Capitol riot in 2021, a retired firefighter from Chicago was harassed by phone calls and attacked on social media after he was misidentified by online crowds [22].

Other researchers are using AI-based facial recognition to identify people in photographs. This technology allows users to quickly narrow down a reference database of potentially thousands of images to an ordered subset of highly similar candidates to the unknown face. As with crowdsourcing, applications of facial recognition technology have yielded both success stories and failures. For example, in 2020, Las Vegas police used facial recognition technology to identify and arrest a murderer [8], and Washington Dulles International Airport caught three imposters in 40 days by using facial recognition technology [3]. Although facial recognition helps to narrow down the candidate pool and can make needle-in-the-haystack identifications faster or more tractable, there are still many false positives in search results that can lead to misidentifications, especially in modern photos. In 2019, for example, police using facial recognition misidentified an innocent black man who was jailed for 11 days [5].

To help close the gap between facial similarity and person identification, researchers have begun to combine the complementary strengths of crowdsourced human intelligence and AI-based face recognition. This approach helps investigators to identify people in photographs by incorporating holistic comparison and biographical information. In prior work, [30] developed a web-based platform, Civil War Photo Sleuth (CWPS), combining both crowdsourcing and facial recognition technology to identify American Civil War-era soldier photos. They found that users were able to successfully identify hundreds of unknown photos using CWPS's "haystack model" person identification pipeline. Despite these successes, the hybrid crowd–AI pipeline also produced some false positives and misidentified photos. Especially when users work with a set of similar faces, it is difficult for users to decide among a subset of promising high-similarity candidates. This challenge is known as the "last-mile problem" of facial recognition, similar to problems in transportation and telecommunications where it is relatively easy to get information or materials most of the way to its destination, but the logistics of the last mile (e.g., getting from the airport back home) are complex, ineffi-

cient, and hard to scale up. As one solution, researchers noticed that users posted potential matches from CWPS on social media and sought second opinions from the Civil War photo community [30, 31].

Given this challenge, Mohanty et al. [29] developed a follow-up system, Second Opinion, aimed at directly supporting the last mile of person identification. In this tool, the researchers combined experts and crowdsourcing and use the workflow of "seed, gather, and analyze" to help users to focus more in details on the Civil War soldier faces by reviewing the unique facial features and others' feedback in a structured and organized way. According to the evaluation, crowds filtered out 75% of face recognition's false positives, and maintained 100% recall rate. However, Second Opinion also had a number of limitations [29]. First, the software automatically selected the shortlist composed of the 5 candidates with the highest face recognition scores; users could not change the size of the shortlist nor modify its contents. Second, Second Opinion recruited paid workers from Amazon Mechanical Turk to serve as the crowd. Participants voiced that they preferred to consult personal contacts and people knowledgeable about Civil War photography, in addition to strangers, when seeking second opinions. Third, Second Opinion focused only on labeling facial feature similarities and differences; it did not allow for detailed discussions or integration of external information sources.

In our work, we have built a new system, SleuthTalk, which adopts the advantages of the Second Opinion system while seeking to address some of its limitations, towards the goal of helping with the last-mile problem in historical person identification. In this new system, we first added multiple user interaction options to allow users to weigh in and pick the potential candidates for a shortlist. Users also have the freedom to add a new candidate or archive it to better manage their shortlist without having the restriction of using the top AI matches. We also designed a private project workspace for users to invite trusted members,

such as their friends or family, to work together and identify the unknown Civil War photos. This private page provide a space for users to share with each other their opinions and pool their individual knowledge to get confirmation, but also help to prevent the spreading of misidentifications during the process. Extending the work from Second Opinion [29], we designed a structured feedback section to allow users to compare each candidate in detail by looking into the facial features. However, we extend this by also providing a space for them to share the other sources outside of CWPS and resolve the conflicting opinions. Using a poll system, project creators can collect and organize members' decisions and determine the final identity as the winning candidate.

We evaluated the SleuthTalk through a mixed-methods study where 6 participants spent one week identifying a Civil War photo on the SleuthTalk and one week to identify a different unknown Civil War photo using existing practices, i.e., a Facebook group. Through the study, we seek to answer the following research questions:

- **RQ 1:** What system features allow for a software tool to address the last-mile problem of historical person identification?
- **RQ 2:** How effective are these features at allowing users to address the last-mile problem of historical person identification?

# Chapter 2

# Review of Literature

## 2.1 Hybrid Crowd–AI Approaches to Person Identification

Artificial intelligence offers great potential in many different fields in our life. However, the results are usually not perfect. Likewise, crowdsourced human intelligence and human computation can solve many problems that computers cannot yet, but these approaches also have limitations. Researchers started to use the hybrid crowd–AI approach to leverage the complementary strengths of both artificial intelligence and the wisdom of crowds in various application areas. First, researchers used the human-in-the-loop approach to improve the machine learning model's performance. For example, Cheng and Bernstein [11] presented hybrid crowd–machine learning classifiers to improve both algorithm performance and human judgment. Zhang et al. [42] proposed a crowd–AI hybrid system to ask crowds to tune and improve the black-box AI algorithm in deep learning-based damage assessment applications. On the other hand, some researchers use AI to assist the crowds' work to ensure a better user experience and results. For example, Guo et al. [19] developed a full-stack crowd–AI camera-based sensing system, Zensors++, to solve the limitations of smart appliances with built-in cameras. Chau et al. [10] used machine learning and visualization to help users to explore large network data and make sense of it. Even though there are a growing number of

hybrid crowd–AI systems in the research community (e.g., CrowdScape [36], PANDA [15], etc.,) there is a lack of research on using the hybrid crowd–AI approach to improve facial recognition or person identification, two very common tasks for which AI is used nowadays.

In this work, we focus specifically on historical person identification, which has seen even less research attention. Yet, identifying people in historical photographs is very important, and has significant value both economically [14] and historically and culturally [7]. Many individuals in photographs are still unknown nowadays, and various experts ranging from genealogists to historians to antiques dealers seek to identify them. For example, of all the portraits of soldiers which were taken during the American Civil War (1861–1865), only 10–20% are currently identified while over 4 million of them are believed to have survived [30].

Because of the low identified rate and high value of the historical photos, various online groups have formed to discuss and try to identify them, including several Facebook groups [29]. While not specifically focused on identifying Civil War photos, two popular genealogy online communities, Find-a-Grave [6] and Ancestry [2], allow users to collect, analyze, and discuss historical person identification, including related documents such as photos, military records, and other priamry sources. However, Willever-Farr and Forte [39] raised a concern about the misinformation in these two communities by allowing inexperienced users to provide information and try to verify the identity. Additionally, they note that AI capabilities such as Ancestry's "hint" leaf icons make automatic recommendations of information to users which may lead them down the wrong path. This study illustrates the socio-technical challenge of historical person identification in which both users and AI can introduce misidentifications.

In 2019, researchers Mohanty et al. built the free public website, Civil War Photo Sleuth, combining both human expertise and facial recognition technology to help users specifically identify American Civil War portraits [30]. This hybrid crowd–AI approach and the

"haystack model", where users "build a haystack" (collect reference images), "narrow the haystack" (reduce false positives with facial recognition and search filters), and "find the needle" (using visual comparison and biographical records), successfully helped identify thousands of historical Civil War photos. But it also brought another concern that the photo can easily be misidentified when users have a set of potential matches which are very similar in both facial features and biographical details. Following this work, researchers Mohanty et al. built another software tool, Second Opinion [29], which uses the "seed-gather-analyze" model to assist experts in solving this last-mile problem when users can not decide the correct match from a set of similar photos. The core idea is allowing users to ask for a "second opinion" from other users to analyze the shortlist. However, as mentioned above, Second Opinion provided a limited experience where users could not control which users helped them or even which photos composed the shortlist, and deeper discussion or incorporating external sources was not supported.

SleuthTalk builds on this idea, not only allowing users to look into facial features on each photo and ask other users of their choice for a second opinion, but also allows users to shortlist their own candidates from facial recognition results. In addition, we designed SleuthTalk to focus not only on facial similarity, but also encouraging users to discuss other external sources.

## 2.2 Collaborative Tasks in Private Workspaces

Many studies have showed that a collaborative workspace helps prevent confirmation bias and helps to solve the problem of limited individual knowledge [12, 13, 38]. Schwind and Buder [37] conducted two studies to find when preference-inconsistent recommendations are effective; the results showed that reference-inconsistent recommendations help reducing

confirmation bias under cooperation and competition. Parmigiani [32] studied the difference in teachers' decision-making as an individual or within a group and showed that decision-making in a group could help reduce the confirmation bias. Kittur et al. [23] also found that sensemaking of large-scale information will be necessary to overcome individuals' limited and fixed individual cognitive capacities for learning, memory, and insight. Collaborative workspaces provide an environment for users to perform collaborative sensemaking, provide diverse perspectives, and review a large amount of information.

However, collaborative work can also have downsides. For example, in a Facebook public group study, Pi et al. [34] showed that online reputation affects users' knowledge sharing attitudes. Users are very careful to share their opinions because of fear of hurting their reputation in a group, a phenomenon known as evaluation apprehension [28]. Furthermore, if the discussion is happening in a larger group, misinformation can be spread quickly during the conversation to outside of the group, where it loses context. For instance, in the public discussion concerning COVID-19 on Twitter, Huang and Carley [21] found that disinformation storylines are more likely to be spread by regular users, rather than malicious actors. This is because many times, users did not know whether it is true when they see a piece of information during the discussion process. Many users would not spend some time investigating and verifying the information before they share it with others [18].

SleuthTalk is helping to tackle these problems by providing project creators with control over user access and invitations to a collaborative private workspace. The speculation and errors that could lead to misinformation spread made by any member are private in the group, but the collaboration enables the benefits of diverse perspectives and processing a large amount of information. The system will only share the final result with the public.

## 2.3 Structured Feedback

Users have many ways to provide feedback or second opinion, but structured feedback offers support for users to contribute more meaningfully [9]. Here, we organized the prior studies to collect structured feedback from crowd workers into either quantitative data or qualitative data. For example, in the design of user interfaces to collect quantitative data, ConsensUS [27] provided a slider for users to share their opinion. The system shows every user's opinion on the same slider with a different color. In addition, this tool starts from capturing independent opinions and then represents the group's opinions to every participant. SleuthTalk adopts this idea to capture each users' opinion on each candidate, then have users vote across candidates to represent the project group's final opinion (displayed as a person icon), which can also be sorted in the order of the number of votes. Another system, Voyant [41], gives users a place to receive non-expert crowd feedback on designs, uses seven different degree options (strongly / moderately / slightly disagree / agree and neutral) for users to vote on their opinion, and then aggregates the result in a vertical histogram. Similar to this design, SleuthTalk also provides five different degree options (highly / slightly confident yes / no and not sure) for users to vote on how much they believe the mystery photo is the matched candidate. Then, the system aggregates the result using a horizontal histogram. Furthermore, Luther et al. [28] presented the web-based system CrowdCrit, dividing the critique into seven micro-tasks reflecting seven design principles (Readability, Layout, Balance, Simplicity, Emphasis, Consistency, and Appropriateness) and allows users to annotate the design itself to indicate the relevant area of the design for that critique. SleuthTalk also adopting this design idea to allow users to crop the areas on the photo to indicate the selected facial feature. At the same time, we separated the tasks into facial similarity comparison, external sources sharing, and final stage voting to help users track their process in the historical person identification task. Unlike all of these systems, which

primarily focus on design critique, SleuthTalk adapts these ideas for the novel context of facial analysis historical person identification.

Researchers have also designed structured ways to collect and visualized crowdsourced qualitative feedback [16, 40, 43]. For example, Kriplean et al. [25] presented a novel platform, ConsiderIt, which asked every user to provide their opinion on either advantages and disadvantages of a topic (e.g., a proposed government policy), which will be displayed in the margin. Then, each participant can organize others' comments by adding to their pros and cons list. Voyant [41], also mentioned above, provided a word cloud of free-text comments for users to choose from and a textbox to express their opinion. SolutionChat [26] designed a real-time message recommendation UI with structured prompts to assist users in sharing their opinions during a community discussion. SleuthTalk builds on top of these ideas in the novel context of historical person identification, not only having multiple micro-tasks which help to collect users' quantitative and qualitative data, but also allowing users to have a final step to vote across each candidate to finish the project.

# Chapter 3

# System Description

We developed SleuthTalk, an online tool as part of CWPS, for supporting users to collaborate with other users for identifying unknown photos in a private workspace. Collaboration helps reduce the possibility of misidentifications that might result from a single user's lack of knowledge or perspective. SleuthTalk allows users to shortlist potential candidates, invite collaborators and discuss with them, compare the unknown photo with the potential candidates, and conduct polls to finalize a winning candidate, if any. In this section, we will describe the three main components of SleuthTalk: 1) shortlisting potential candidates, 2) private collaborative workspace, and 3) structured feedback.

## 3.1 Design Alternatives

We explored many different design iterations while creating the workspace structure and layout. For example, in one early design (Figure 3.1), we displayed the candidates' thumbnails (faces) on the left side, with mouse hovering to manage the candidates, compare the facial similarity, view the details of the biographical information, and show voting results. On the right side, we displayed the unknown photo's detail, discussion, and poll. Users could also manage the project, including the project members, in the settings.

However, we did not adopt this design because of three main reasons. First, we only emphasized candidates' thumbnails (faces), but our focus was also on the external sources beyond

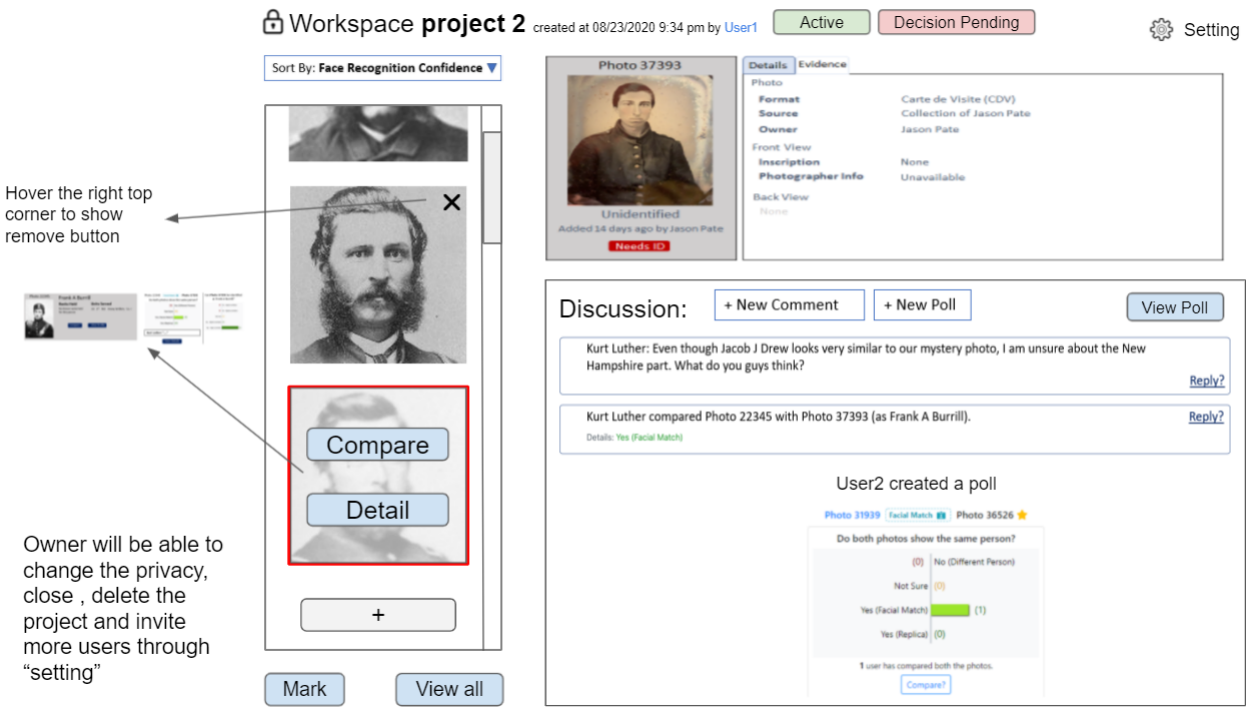


Figure 3.1: Previous Workspace Structure Design.

facial similarity, enabling a holistic comparison. Second, much information is hidden in this design, and users need to take an extra step to find where they can compare, view the details, vote in the poll, or manage the invitations. Thus, we were looking for a much simpler and clear design. Third, we integrated poll vote into the discussion section and used a modal dialog box to display each candidate in a horizontal carousel (Figure 3.2). However, this design contrasts with our original idea to let users vote in one poll across every candidate. We sought a new design to provide users enough information reflecting their comparison and identification votes, discussion, and the candidates' detail. In particular, users should be able to vote in the poll across different candidates in the shortlist instead of looking at them one at a time.

After many sessions of adjusting, brainstorming, and discussion, we achieved the current design of the workspace. In the following sections, we describe in more detail how we


Figure 3.2: Previous Poll Vote Design.

structured each component in the SleuthTalk to address the last-mile problem of historical person identification.

## 3.2 Shortlisting Potential Candidates

### 3.2.1 Shortlist the candidates and create a project

After the user uploads a new unknown Civil War photo, the system will retrieve all the facially similar results in a descending order based on the similarity score (confidence score) calculated by the Microsoft Cognitive Services Facial Recognition API [1] [30] while also satisfying the search filters. Users can carefully inspect individual search results to determine whether they are facially similar to the query photo and whether the biographical information (e.g., military service records) line up with the visual clues of the query photo (e.g., uniform insignia). However, prior work has shown that users have sometimes identified the correct match beyond the top-5 and top-50 face recognition results [30]. Similarly, benchmarking studies of CWPS have shown that the top match is not always the correct match; for example,

[1] https://azure.microsoft.com/en-us/services/cognitive-services/face/

one confirmed match was ranked 66 and another match was ranked 657 in the search results list [31]. SleuthTalk is designed in a way to allow users add any potential matches to a shortlist from the facial recognition results and create a project for analyzing these shortlisted candidates in a focused manner.

There are two ways users can accomplish these goals(see Figure 3.3). First, when a user runs a search for a query photo, each result contains an "Add to Shortlist" button on hover. Clicking the button causes the candidate to be added to the shortlist; the background of that candidate's search result changes color, and the candidate is added to a floating container on the right side of the screen. A candidate can be removed from the shortlist by clicking the "Remove from Shortlist" button on hover, in the same manner. Once a shortlist is created, the user can create a project by clicking the "Create Project" button either at the top of the search results list, or at the bottom of the floating container.

Second, the user can modify the shortlist once the project is created, as described below.

Shortlists are represented in the database as a unique relationship between a photo and a user. That is, each user can have one shortlist for each photo. Correspondingly, users can have one project per shortlist (i.e., per photo). It follows that multiple users can each have a different shortlist for the same photo, allowing different social networks to explore the same photo mystery from different perspectives.

After the user creates the new project, SleuthTalk will display each selected candidate under the "Shortlisted Candidates" section. We display the biography information and the photo thumbnail (face) in each candidate card (see Figure 3.4). However, if multiple faces link to one candidate, we use the vertical carousel to display all the thumbnails, one at a time, which can be cycled by clicking up and down arrows. Bubble icons indicate the number of available photos (faces) for a candidate. When the user refreshes the project page, all the

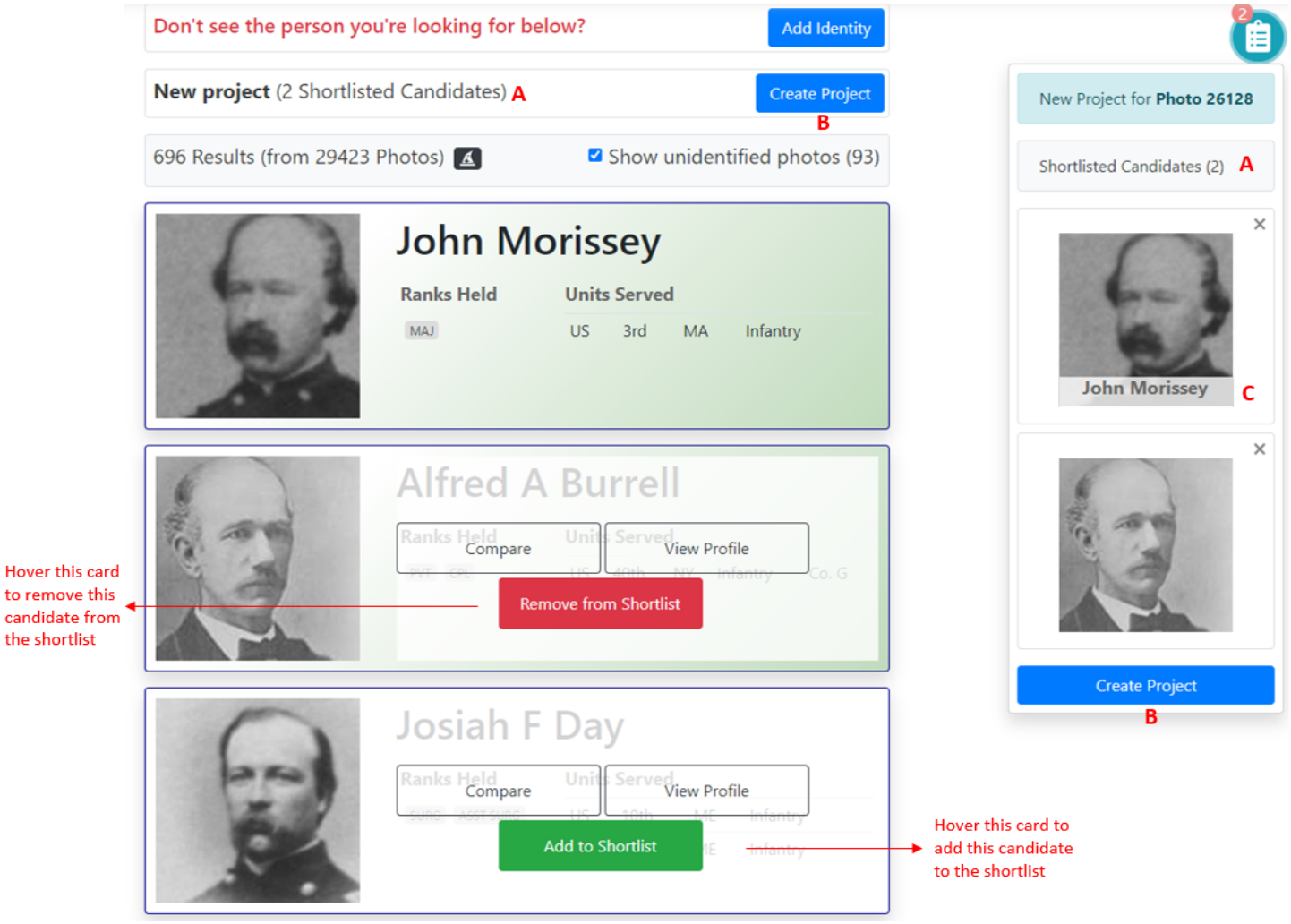


Figure 3.3: Portion of the search result page. We show the total number of candidates that have been selected for the shortlist at position A. The button to create a new project is at position B. Users can hover the face in the container to see the identity at position C.

shortlisted candidates will be randomly sorted to avoid a common bias where people always focus more attention on the first candidate. Moreover, we also provide an option for users to sort the candidates in various popular ways, i.e., by the time they were added, or the number of votes in the poll.

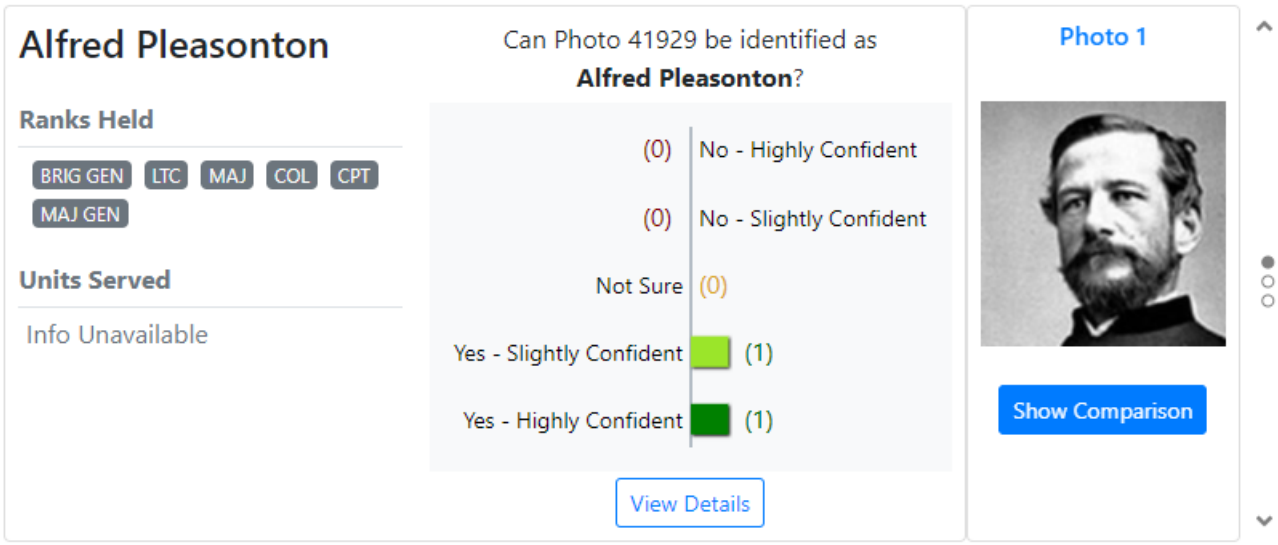


Figure 3.4: Candidate Card.

### 3.2.2 Modify the shortlist

If users want to add more candidates after the project is created, any project member can click on the "Add Candidate" button and follow the same process using the search results page to add a new candidate. And to help the user better manage their shortlist without too much overwhelming data, we designed the workflow to only to allow users to add one candidate each time.

Users can also archive the candidates on the shortlist if they change their minds. However, users can only archive the candidates which they added themselves; the exception is that the project creator can archive any candidates in the shortlist. Even though the candidates can

be archived, this is not permanent. SleuthTalk also displays the previous votes and other information for the archived candidates in the archived section (see Figure 3.5), so any users in the project can always refer back to the previous information and restore an archived candidate from the project page if they need.

It is not possible to entirely delete or remove a candidate. Instead, they can only be archived, so that the aforementioned history and restoration features remain available.

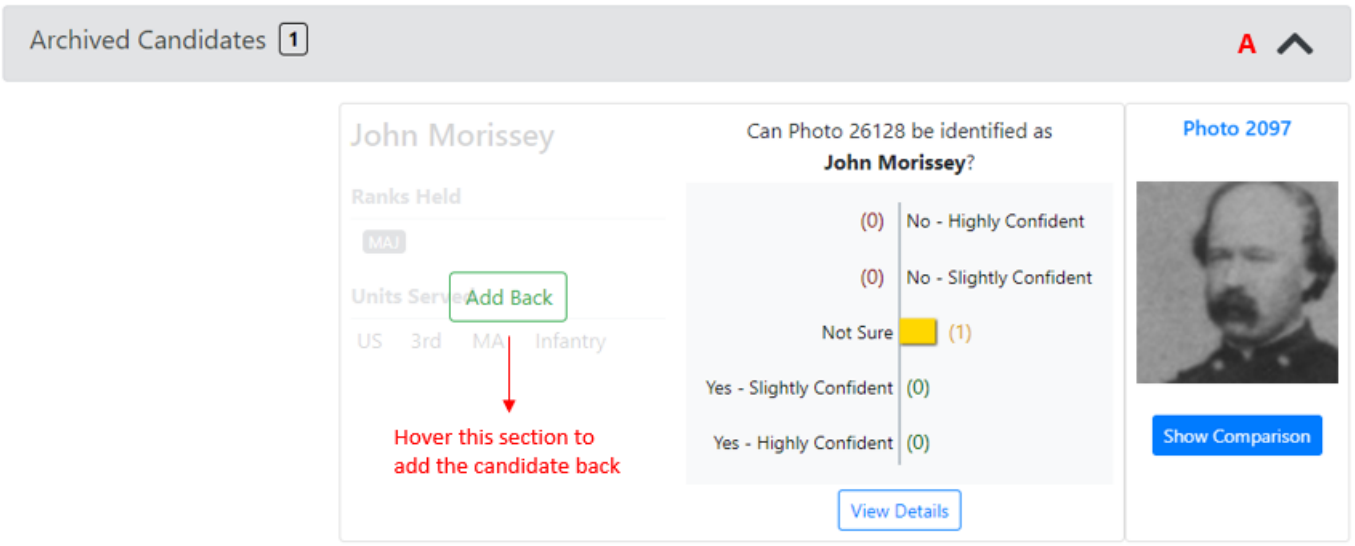


Figure 3.5: Archived section with an archived candidate card. Users can hover over the biography information column to add this candidate back to the shortlist. They can also click the arrow (position A) to expand or collapse this section, which is collapsed by default.

## 3.3 Private Collaborative Workspace

According to previous work [29], users of CWPS stated that it would be helpful if they knew more about the expertise of the person who is providing the second opinion. Moreover, CWPS users are more likely to discuss the identity of the unknown photo with their family and friends. Therefore, we designed SleuthTalk to be a private workspace for them to

invite anyone they would like to and collaborate with them, but otherwise exclude other users from viewing the project and discussion. Additionally, creating a team leverages the benefits of multiple perspectives and can help the project creator avoid mistakes because of the individual's limited knowledge and the possibility of confirmation bias and overlooking the facial comparisons. The project leader can invite up to 9 team members by clicking on the "Add Member" button. This interface allows them to invite CWPS users with existing accounts by typing their name, username or email; an auto-complete feature allows users to quickly find the one they are looking for. Further, users can invite non-CWPS users simply by providing their email address. The invited members will receive an email with the link to participate in a project (see Figures 3.6 and 3.7).

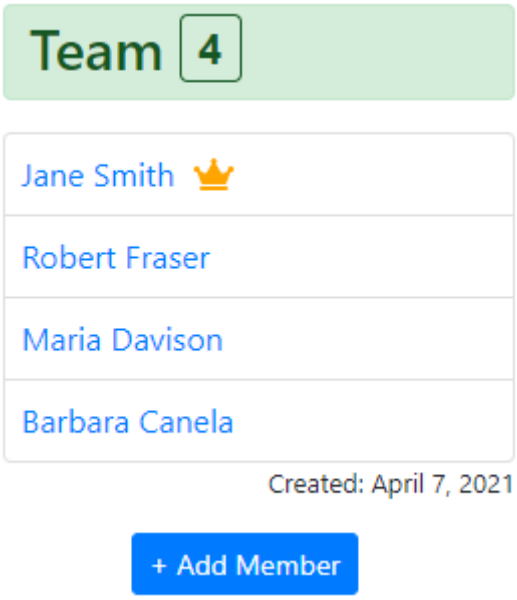


Figure 3.6: Invite Team Member. The first team member with a crown icon above to it is the project creator in the workspace.

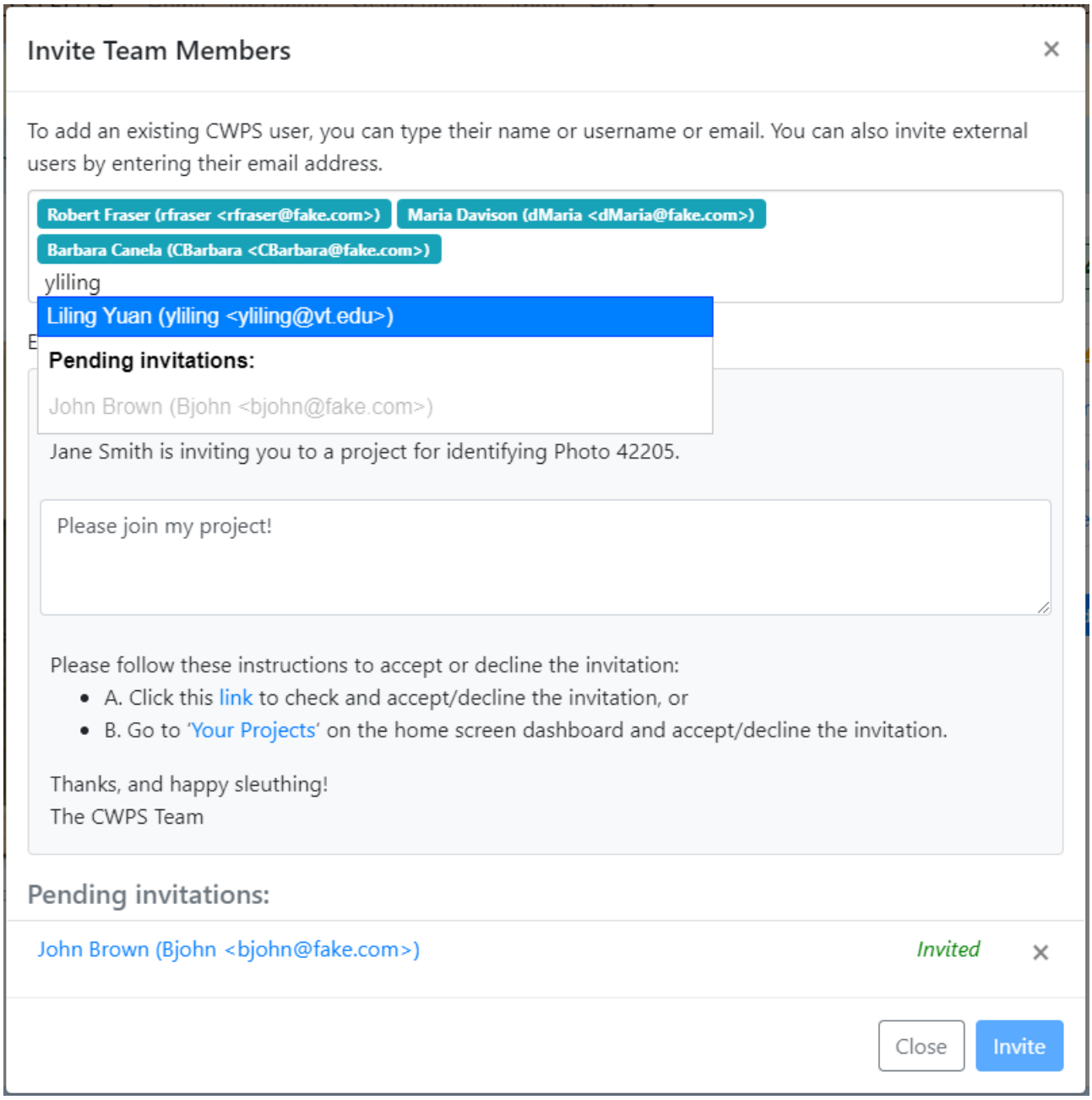


Figure 3.7: Invitation Dialog Box. The project creator can view all the users who have already accepted the invitation or have not yet provided their response. Also, the project creator can customize the invitation email to explain the photo or the project to the invited users.

In SleuthTalk, invited users who have not yet accepted can only view the information on the project page at that time. Once users accept the invitation, they are allowed to interact with other team members on the page and view the rest of the information. Other CWPS users who are not project members cannot view votes, discussion, or other related information on this project page because of privacy settings. When the project creator finalizes the identity as the winning candidate, the identity of the unknown photo finally becomes public information to everyone on CWPS. However, other CWPS users who are not project members can still not view the details of the project page. We use this approach to set up the private collaborative workspace to prevent misidentification during the discussion before the user finalizes the identity; but this "publishing" mechanism provides a way to share the results with the broader public for them of benefit. Simultaneously, we are also helping to create this workspace for the users who would like to keep the identity of the unknown photo and the resources private from others permanently.

## 3.4 Structured Feedback

### 3.4.1 Facial similarity comparison

There is a two-step process to compare two Civil War photos on the current CWPS website. First, users need to compare the two photos based on only facial similarity, and second, they express how confident they are about the photo's identity (ID), incorporating broader context about the proposed ID's biography. In this first step, users only focus on comparing each individual candidate, so they are allowed to have a positive vote for multiple candidates in the shortlist. According to prior work, the Second Opinion system has shown that comparing the facial feature gives users more confidence in making an identity [29]. Therefore, in addition to

the current two steps, we add a third step based on Second Opinion to the process, allowing users to compare fine-grained facial features (see Figure 3.8). In this step, we provide an auto-completed list of 7 default feature examples in the beginning (hair, eyes, nose, ear, facial hair, eyebrow, jawline), and more user-generated options as users provide additional facial features, so that we can avoid having the same facial feature with slightly different names when we display the comparison details [24]. We believe that users may deliberate and carefully compare the two photos for facial similarity by completing this step.

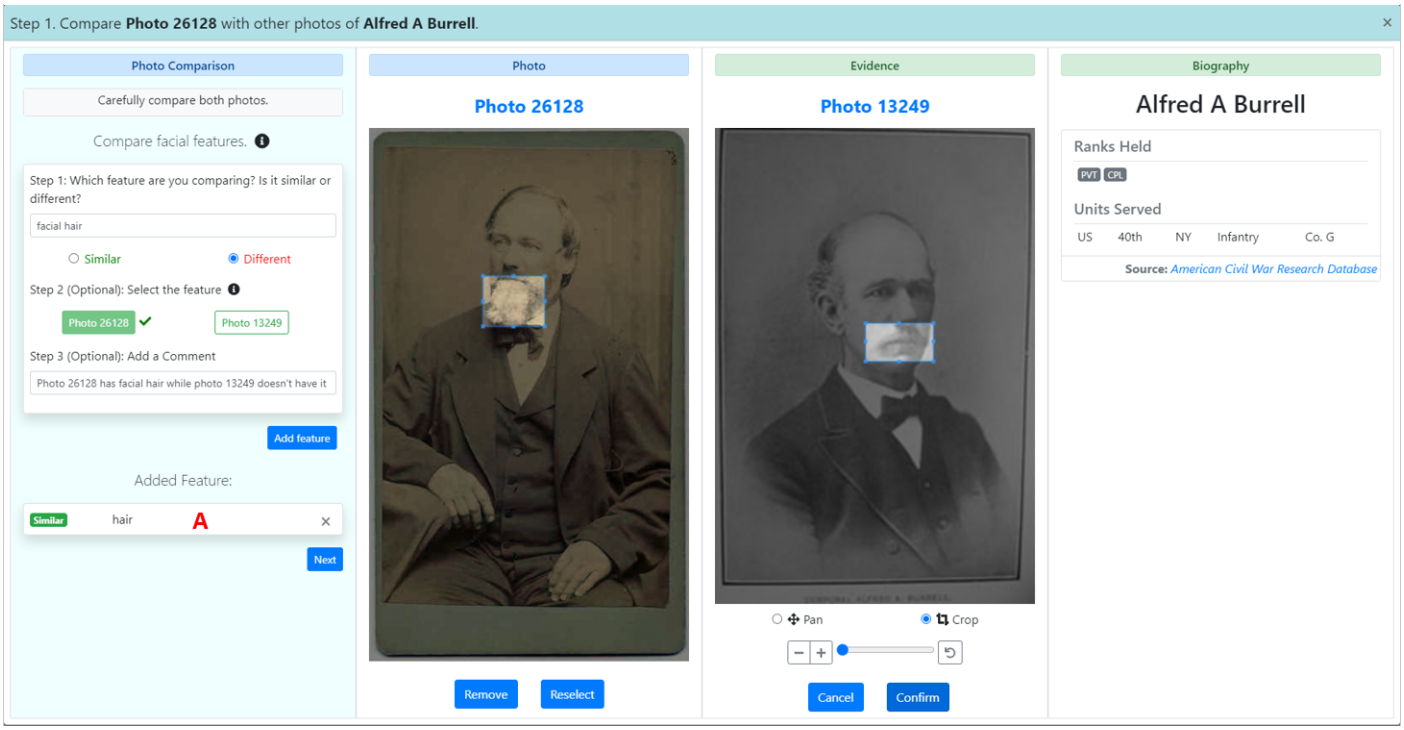


Figure 3.8: Step 1: Facial Feature selection. User can reactivate the cropping tool after they confirm and re-select it again on the photo. User can also hover over section A to see the user's comment in the tooltip and the cropped facial feature appears on the photo.

After the user has compared each candidate and provided a vote, we build on existing designs from CWPS to use a horizontal histogram to visualize the results about a user's confidence in an ID (see Figure 3.9) and the information about how two photos are similar according to both the community and face recognition (see Figure 3.10).

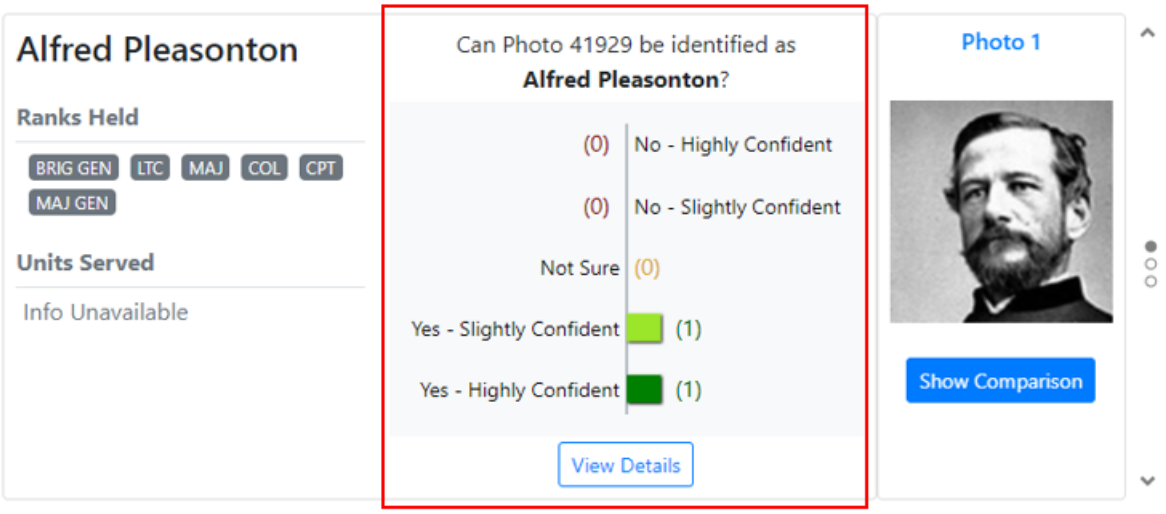


Figure 3.9: Identification Vote in Candidate Card.

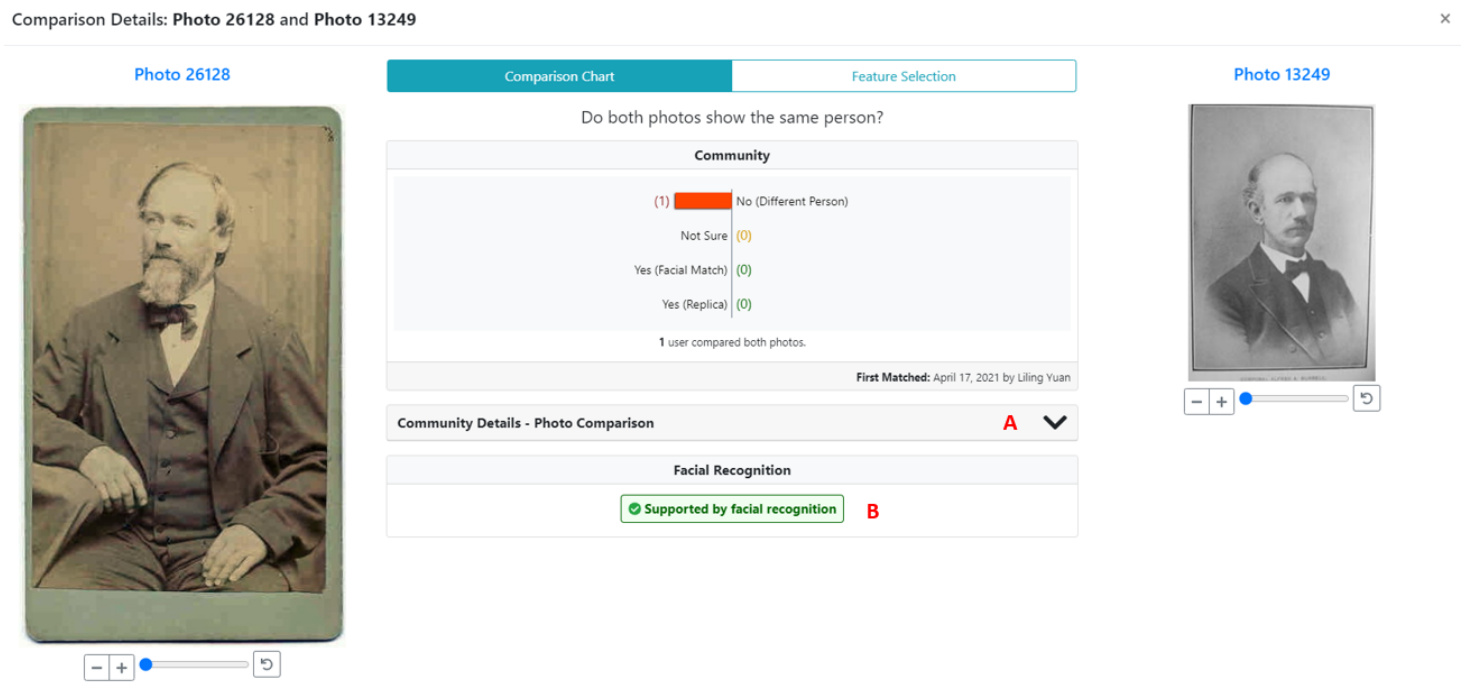


Figure 3.10: Facial Comparison Chart. User can click section A to see the comparison vote details. SleuthTalk shows whether the two photos are supported by facial recognition in section B.

In addition, SleuthTalk displays the facial features that every project member selected during the comparison process, including the name of the facial feature, similarity, cropping box, and comments (see Figure 3.11). We believe that displaying the facial features that other project members selected can help users to narrow down the shortlisted candidates.

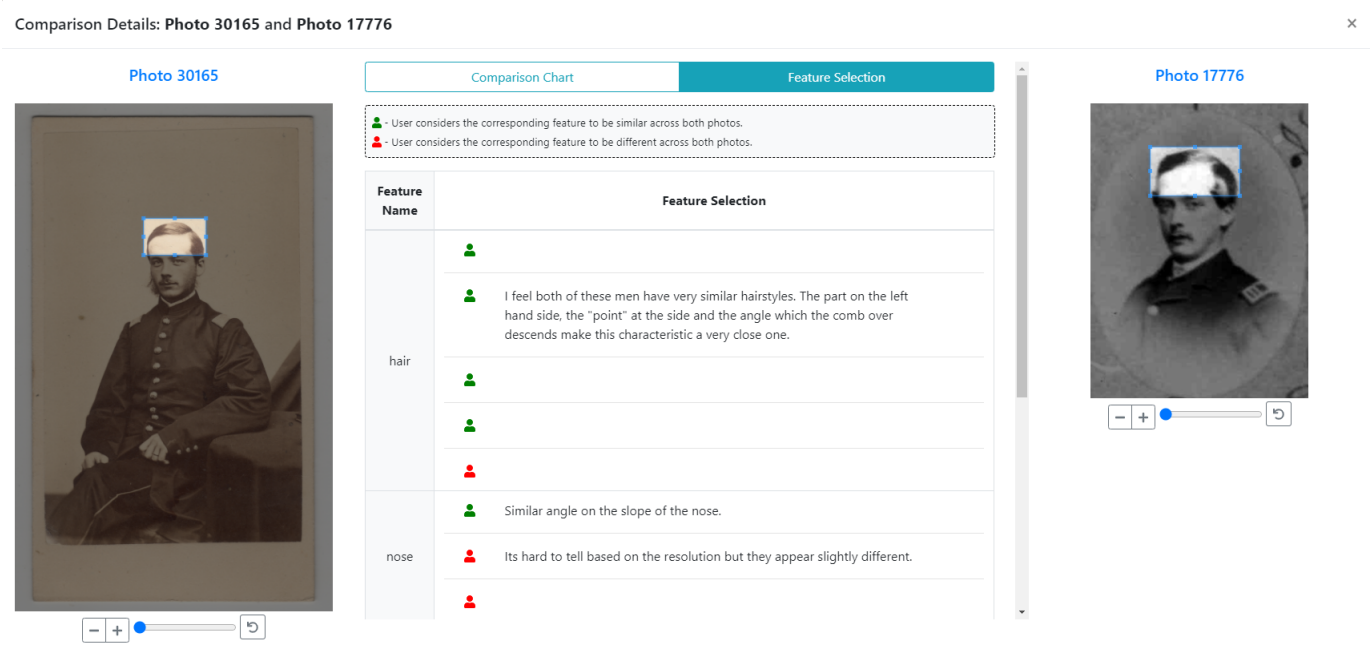


Figure 3.11: Facial Feature Selection Interface. User can hover the selection to view the cropping detail.

### 3.4.2 Discussion Support

Even though we designed SleuthTalk to help the user easily view other people's opinions on facial similarity for each candidate, we also add a discussion section to allow users to have more detailed open-ended conversation, share information sources, and resolve conflicting opinions. Users can also refer to specific photo comparisons via hashtags. For example, if the user wants to see the comparison vote and the facial feature selections between the unknown photo and photo 13249, users can put hashtag "#13249" in the discussion thread

to help others to quickly view the information in Figures 3.10 and 3.11 as a modal dialog box (see Figure 3.12).

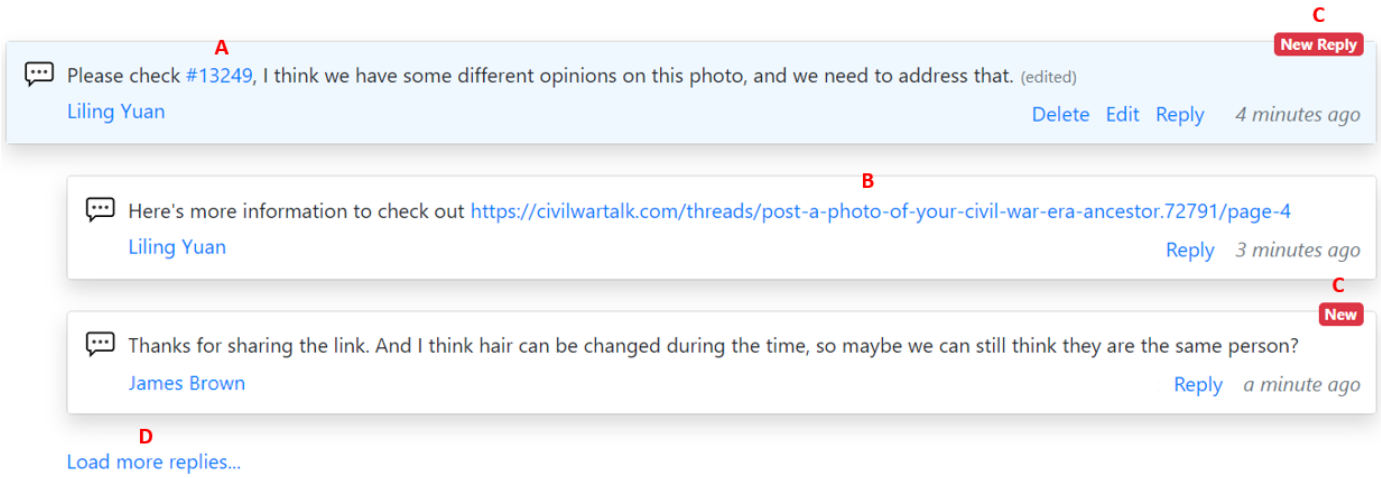


Figure 3.12: Discussion Thread. Position A is an example of the hashtag; users can click to view the face comparison vote detail between the unknown photo and photo 13249. Position B is a clickable link to help user to open external links and information sources conveniently. Position C is a red badge which indicates a new comment or the new thread discussion. Position D is the link where users can click to load more replies for this thread.

In this section, to help avoid redundant discussion and provide a faster way to refer to the shortlisted candidates, SleuthTalk displays all the thread discussions from old to new to encourage users to read previous conversations before starting a new one. At the same time, we initially display the first two comments of each thread to provide more information and encourage users to join in the discussion. As explained in Figure 3.12, the system highlights new replies with a red badge to draw attention to them.

### 3.4.3 Identification Poll and Final Decision

Finally, we designed SleuthTalk's identification poll section to help users collect and visualize one other's opinions on the final identity of the query photo. Any members who finishes comparing all the candidates can then start the identification poll. Then SleuthTalk displays

the poll section only to the users who finished comparing, so that we can show the task step-by-step to avoid any confusion for the users who are still working on comparing the candidates. All the poll votes will be displayed on the left side of each candidates' card. Users can only vote for a maximum of one candidate in the poll section, but we also provide an option for the user to vote on none of the candidates if they could not find a strong match in the shortlist. SleuthTalk requires at least half of the project members to vote on the candidate to proceed to the next step.

Project members vote in the identification poll based on the current shortlist. If the shortlist is changed by a member adding a new candidate or archiving a candidate, the poll is automatically reset (see Figure 3.13).

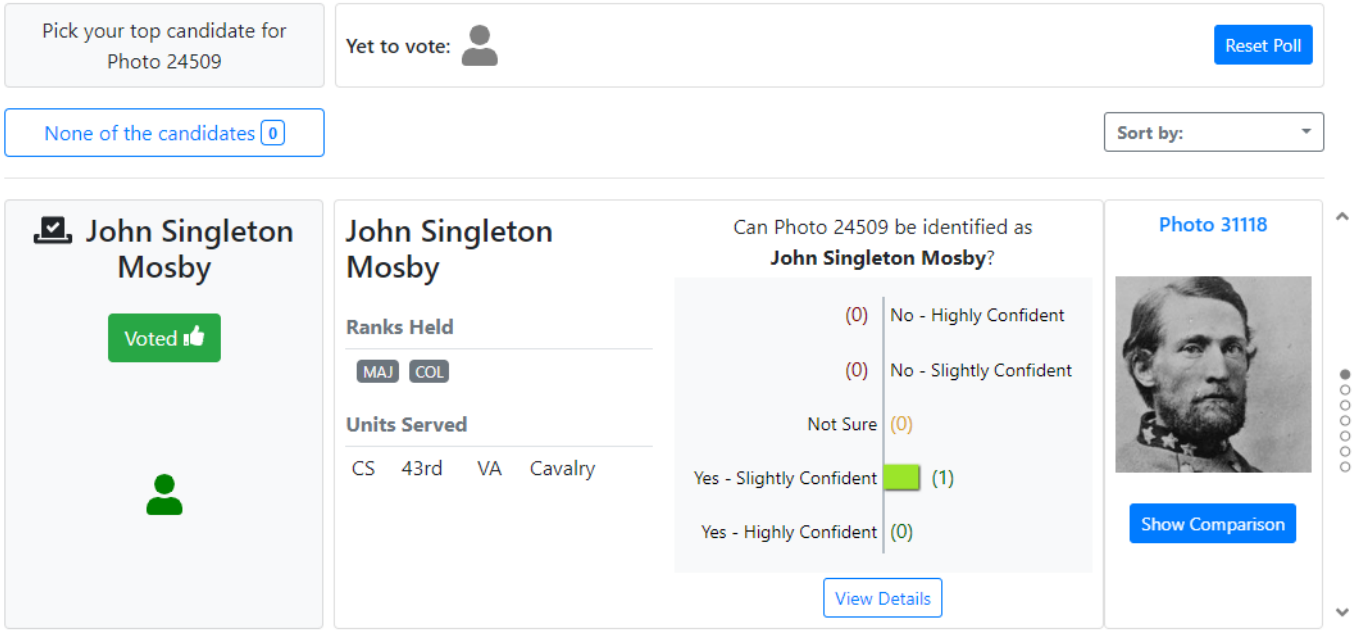


Figure 3.13: Identification Poll Vote Section. We use the person icons to help project members to track the voting status. They can hover the icon to see the voter's name. Members who have not yet voted are represented at the top as "yet to vote."

Once at least half of the project members vote on the candidates, SleuthTalk displays further instructions. Two outlier situations are possible. In the event that the vote is a tie, the system

then shows a yellow warning alert on the top to indicate this situation. There is also the possibility that majority of members vote on none of the candidates; in this case, the system shows a red alert on the top. If there is no tie and we have a winning candidate, the system displays a button for the project creator to finalize the identity of the unknown photo as the winning candidate, and share the identity with the public in CWPS.

## 3.5 Implementation Details

SleuthTalk is built upon the web application Civil War Photo Sleuth [30] using Javascript, HTML, CSS for front-end interfaces and Python and Django for back-end processes. We used the PostgreSQL database to store the data and Amazon S3 to store the images. On the project page, we also adopt WebSocket and Redis to achieve real-time discussion and interaction with the use of the ASGI (Asynchronous Server Gateway Interface) server.

# Chapter 4

# Evaluation

We conducted an exploratory, mixed-methods evaluation study to understand how well SleuthTalk supports users in collaborating with other users and giving structured feedback for identifying unknown Civil War photos in a private workspace.

## 4.1 Recruitment and Participants

We shared a video demonstrating the different features of SleuthTalk on CWPS and different Civil War photography pages and groups on Facebook for recruiting interested participants. We recruited six participants with diverse experience levels in identifying Civil War photos (mean = 16 years, min = 0, max = 44). Out of the six, five identified as male and one female, and were distributed across different age groups (min range = 18–20 years old, max range = 51–60 years old). Each participant received a $50 Amazon gift card. We anonymize the participants with the identifiers P1–P6.

## 4.2 Procedure

All six participants received access to SleuthTalk on CWPS. In order to understand how SleuthTalk compares against a relevant baseline, we chose a popular private Facebook group

with over 13,000 members where users currently seek help identifying Civil War portraits.[1] Each participant was assigned the task of identifying one unknown photo using SleuthTalk and another using the Facebook group. In the first week, P1–P3 were assigned SleuthTalk, while P4–P6 were assigned to Facebook. In the second week, they were interchanged.

Participants were free to invite anyone (existing CWPS user or non-user) to be a part of their project team on SleuthTalk. In addition to that, we sought consent from our participants and shared their CWPS usernames with the other participants, giving them the option of inviting each other but without any requirement to accept the invitations.

For the SleuthTalk task, participants were instructed to upload at least one unknown Civil War photo and create one project (per photo) with candidates shortlisted from the CWPS search results page. They were given the option of inviting other members for their project, without any restriction on who or how many. Participants were given one week to use the features available to them for identifying the photo and wrap up the project (i.e., reach a decision by conducting the final poll).

For the Facebook task, participants had to post a different unknown photo to the Facebook group and seek help from the group members for identifying the photo. Since the group is private, we did not directly observe or collect data related to these online interactions; instead, we asked participants during our interviews to recall their experience for us.

## 4.3 Data Analysis

We collected and examined the log data for each project for understanding how users used the SleuthTalk system. We specifically analyzed the different team compositions, their discussions and how they compared the different shortlisted candidates to the unknown

[1] We anonymize the Facebook group to protect the privacy of its members.

photo. We further analyzed the final polling data for each project.

## 4.4 User Interviews

We also conducted in-depth, semi-structured interviews with all 6 participants about their experience with both SleuthTalk and Facebook. All interviews were conducted on Zoom and were recorded, fully transcribed and analyzed with respect to the following themes: shortlisting process, collaboration, structured feedback, and privacy.

# Chapter 5

# Findings

## 5.1 Overview of Project Usage

According to our log analysis, we found that there were 9 projects with 9 different unknown Civil War photos that 6 participants created during the study (3 participants created 2 projects each). The photos they uploaded were either from their family, the auctions, or purchased from friends. Although 5 participants tried to identify the photo before the study, they did not get ideal results; i.e., they were unsuccessful in securing an identification they were confident in.

Across all nine projects, participants sent 45 invitations to join their projects, including duplicate users from different projects, encompassing both study participants and the project creator's friends. Of the 7 users who are not among the study participants, 3 of them (2 existing CWPS users and 1 new user) accepted the invitation, and the remaining 4 people (3 existing CWPS users and 4 new users) did not respond before the study ended.

Looking into the shortlists that participants created, participants added a total of 37 candidates initially (mean = 4.11 candidates per project), 1 candidate was added after the project was created, and no candidates were archived during the study. Examining the comparison vote between each candidate and the unknown photo, we saw that, across the 9 projects, a total of 254 facial features were selected, and 89% of them use the predefined feature names.

Even though all the participants mentioned in the interview that facial similarity is one of the critical reasons they chose to add the candidates to the shortlist, around 57% of all the selected facial features are labeled 'different' between the candidate and the unknown photo.

At the end of the study, 7 out of the 9 projects started a final identification poll, but none of the project creators made the final decision on the unknown photo's identity using the software. Participants cast 125 identification votes during the polls, but only 8 of these were positive.

Comparing the SleuthTalk results to the Facebook task, half of the participants self-reported that they determined their uploaded photo's identity with the help of members of the Facebook group.

## 5.2 Project Profiles

In this section, to provide context, we present summarized profiles of each of the 9 projects created by participants during the study.

### 5.2.1 Project 1

P2 created two projects during the study, Project 1 and Project 2. In this project, P2 invited 3 users (2 other participants and 1 friend who was new to CWPS), but only 2 of them (1 participant and the 1 friend) accepted the invitation. P2 added 2 photo candidates to the shortlist. Both received negative votes from everyone in the project. There was only one vote (none of the candidates) in the poll started by a project member. Therefore, there is currently no sign that P2 will determine the identity of the photo he uploaded and be able to close this project.

### 5.2.2 Project 2

Project 2 is the second project that P2 created. This project did not see much user activity. In this project, he only invited one user (his friend who is new to CWPS) to identify the photo, and his friend accepted the invitation to join. Even though P2 added three candidates to the shortlist, they all got negative votes from both P2 and his friend. Furthermore, there is no other action to proceed with the identification process, such as modifying the shortlist or discussing and sharing other sources outside of the CWPS.

### 5.2.3 Project 3

Project 3 was the largest project in the study. Created by P4, it had 7 people who contributed to the identification process. Even though P4 invited three of his friends (2 existing CWPS users and one who never used CWPS before), only one existing CWPS user joined the project. This project contains 4 positive votes of 29 total votes from participants across 5 different candidates in the shortlist. This includes two polarized votes for two candidates who receive both high confidence no and high confidence yes. In the end, project creator P4 started the poll and received 3 votes from 3 members in the project, supporting various candidates in the shortlist as the winning candidate. Because there are not yet enough votes from the project members, this project still remains open.

### 5.2.4 Project 4

The fourth project was also created by P4 with the same invited members as Project 3. All the members of Project 4 were participants in the study. In this project, P4 added 3 different candidates to the shortlist. As the most controversial project in this study, 4 out of

6 project members voted on the same candidate in the poll, although they voted for 'not sure' during the identification vote process. Even though this is the only project which satisfies the requirement to determine the unknown photo's identity, project creator P4 still left this project open and kept the identity of the photo he uploaded officially unidentified.

### 5.2.5 Project 5

Project 5, created by P3, is the project that contains the most discussion. In addition to the project creator, there are 5 other members in this project, including one existing CWPS user who is not among the study participants. In the shortlist for this project, one of the candidates received a positive vote from a user, but this user voted for 'none of the candidates' in the poll later. In the discussion section, one user restated the opinion about the facial feature after providing 10 facial feature selections across three different candidates in the shortlist. But the remaining 3 discussions are all about sharing people's opinions on the candidates they are not sure about. Because more than half of the members voted for 'none of the candidates', this project was left open.

### 5.2.6 Project 6

P5 created Project 6, which includes 2 other study participants in the project. Even though P5 added three different candidates to the shortlist, they all received negative votes and all the facial features were labeled as 'different'. The poll was never started in this project, and this project was left open for more future work.

### 5.2.7 Project 7

Project 7, created by P1, includes all the study participants. Additionally, P1 invited two outside users who never used CWPS before, but neither accepted the invitation. According to the logs, we observed that P1 added a candidate based on the pre-identify search. However, the middle name that P1 searched for is different from the middle name that P1 added to the shortlist. For all three candidates in the shortlist, none of them received any positive votes. Therefore, it is easy to understand why the project members only voted for 'none of the candidates' in the poll. Because the majority of the members did not pick any candidate, this project remains open.

### 5.2.8 Project 8

Project 8 is one of the two projects created by P6. In this project, he invited 5 other study participants, but only 3 of them accepted the invitation. Looking at all the identification votes, we did not see any positive votes on any candidates in the shortlist. However, in the poll, two project members voted for two different candidates they are not sure about and shared their opinion in the identification vote. Because this project has a tie and there is no discussion related to other biography information, this project also remains open.

### 5.2.9 Project 9

Finally, Project 9 was also created by P6. He invited all study participants to the project as well, but only 2 of them accepted the invitation. Of the 5 candidates in the shortlist, one was added after the project was created. Two of them received a positive vote from the user. One user shared his reason to vote for 'slightly yes' by considering both facial similarity and

the inscription on the photo on one candidate in the discussion section, but he also suggested that more examination is needed. However, at the end of the study, the poll in this project was never started, so this project also remains open for more investigation.

In the rest of the Finding section, we will report key observations for each major component we designed in SleuthTalk, as well as comparisons participants made between SleuthTalk and Facebook.

## 5.3 Shortlisting Potential Candidates

### 5.3.1 Participants did not restrict themselves to the top AI matches when creating a shortlist.

According to the usage logs, we observed that eight out of 37 shortlisted candidates were not among the top-five AI matches across all nine projects.

Moreover, participants did not just select the candidates among the first dozens of choices. Participants also chose the candidates spread throughout all the results displayed on the search results page, which can encompass hundreds of photos. P4 explained, "*I went through as many photos as I could; I did not just limit myself to the top.*"

During the interviews, participants also mentioned that they did not limit their attention to looking at facial similarity when deciding whether to add a candidate to the shortlist. Instead, they also looked at the broader context, such as the clothing, equipment, name, and biographical details. P1 explained that "*[I] also take into account what matches in [the] uniform.*" Beyond the uniform, some participants stated that they would also look into the geographic region and other biographical information for each Civil War photo

displayed on the search result page. Furthermore, even though CWPS only displays face recognition results for Civil War photos with the similarity score threshold of 0.50 and above, we found that there is still some possibility that participants select shortlist candidates which have similarity scores below 0.50, through the pre-identification process. For example, one participant searched a name and added that candidate to the shortlist even though the middle name that P1 searched for is different from the middle name that P1 added as the shortlist candidate. Based on this finding, P1 explained in the interview, "*the name was part of the reason I added him because [I was] under the impression that [the last name of] the photo's [identity] is Wells.*"

### 5.3.2 Most participants did not add or archive candidates to manage their shortlist.

Based on our logs analysis, we found that only one participant added a new candidate after the project was created. This participant, P6, explained that "*I didn't have enough time when I first went through. I went back through it. Eventually, I saw one photo that might be [the one], and I added [him to the shortlist].*" Other participants, in the interviews, voiced reasons they might add candidates, even if they did not during the study. P3 found another view of one of the candidates in his shortlist, and told us that he would like to add the new photo to his shortlist "*so you could build up your argument a little bit more.*"

No participants archived candidates during the study. One reason may be that the archiving feature is insufficiently discoverable. Another possibility raised by participants is that the study duration was too short to make archiving useful. All 6 participants stated in the interviews that outside of the study, they would treat the project as a long-term effort and would add or archive the candidates for better management when they find more potential

candidates for the unknown photo.

In addition, one participant expressed a desire to show everyone's opinion on archived candidates. P4 believed that there is a chance that other users who join the project later might provide a different opinion than everyone else in the project. In P4's words, "*If in the future, you invite somebody else, somebody new to the project. Then they go through the ones that have already been there, and maybe they might see differently.*" Although SleuthTalk has this feature implemented, no participants archived a candidate, so this feature was never used by anyone during the study. But the participant's comments helps to support that this feature is needed in historical person identification collaborative workspaces.

## 5.4 Private Collaborative Workspace

### 5.4.1 Participants invited a variety of people to the workspace.

According to our interviews and observations, participants invited a variety of users to their workspaces. The first kind was their family member. For example, P3 stated, "*I also invited my wife, and she has her [CWPS] account as well.*" The second kind of invited member was the participant's friend. For example, P2 invited one of his personal friends to help with identifying the unknown photo in both of his projects. The third kind of invited user we observed were experts. For example, P6 invited an expert within the Civil War photo community. In P6's words, "*He does this professionally, and you know so to have him in [the project] is very important.*" Lastly, participants also invited strangers to the project; P6, for example, explained in the interview that he invited someone he never knew before the project.

### 5.4.2 SleuthTalk helps with combating confirmation bias.

Many participants mentioned that they would like to externalize their current thinking by voting on one candidate in the poll even though they were not 100% sure about their answer. In the words of P5, "*I really became conflicted because there's a lot that looks the same. After I considered it a little more, I thought I'm going to put my voice here that I think it's him.*" Similarly, P6 voted for his preference and said, "*I thought that this [candidate] was the best match. I'm just not sure it's a direct match.*"

Although we designed the poll section for users to vote for the best match across all candidates, we envisioned this poll as part of a longer discussion. We hoped participants would find other external sources to provide evidence for their votes, or propose a new candidate during the process of resolving the conflicting opinions between project members. Indeed, Project 5 provided a rich example of how the poll and broader discussion directly combated confirmation bias for the project creator.

In Project 5, we observed that one candidate received one positive vote from the project creator, P3. However, that candidate received 3 negative votes and 2 "not sure" votes from other project members during the identification voting process. At the end of the study, P3 changed his mind and started a discussion thread, where he explained that"while I agree with <project member> that <candidate> seemed like the most promising, after thoroughly comparing all candidates to my image, I have to come to the conclusion that none of them is my officer."

Later, in the interview, P3 told us that he went back to do more research about this candidate after receiving many counter-votes from other project members. He explained, "*I'd also found some other pictures of <candidate> when he was younger, without facial hair.*" P3 changed his mind after comparing with the new photo he found with the unknown photo.

### 5.4.3 SleuthTalk's private projects provide a workspace for users to discuss the identity of the unknown photo, but also help to prevent the spread of misidentifications.

Participants stated that they were being cautious of sharing information and their opinion on the identity of the unknown photo in public. In the words of P5, "*[I try] not to identify someone if I am not 100% sure because it's hard to undo that. And I actually did make that mistake once and it still haunts me.*" P2 also expressed his feelings on misidentified Civil War photos, saying, "*I generally don't like to make something set in stone. This is who it is, unless there's a lot of evidence to support it. 'Cause I think it's kind of bad to misidentify people.*" However, while participants are very careful of sharing misinformation, there are some possibilities that misinformation is unintentionally spread out to the public. For example, in Project 4, which P4 created, four project members voted on one candidate as the winner of the poll, yet this candidate received six identification votes as "not sure". Even though participants explained that they might need more evidence to prove it, they still put their votes to a candidate to express their opinion while not being sure about their answer. However, these participants do not see their vote as confirming the identity; they are just sharing their thoughts on the identity of the unknown photo. This situation occurred for 5 of the study participants across 4 different projects (3, 4, 5 and 8). They all voted in the final poll for a candidate even though they voted 'not sure' in the identification vote.

If these projects were open to the public, other users might only focus on the poll result and accidentally take this candidate as the winning candidate, despite there being a higher chance that this unknown photo has been misidentified. Fortunately, the private workspace in SleuthTalk helped to provide a way for users to express their thoughts without sharing the misinformation during the identification process to the public.

### 5.4.4 The controlled private workspace helped to prevent toxic comments and counterfeiting.

In the interviews, we were surprised to learn that participants faced challenges when seeking Civil War photo information in larger groups on social media such as toxic or distracting comments, and worries of counterfeiting or even theft. Participants expected that the membership control features and smaller size of the private workspace can help mitigate these risks. For example, one participant shared a story of him receiving an offensive message from another user on Facebook when he posted a photo to a Civil War group. P2 explained, "*Oftentimes it is a very toxic community on Civil War forums, where they are like, [they] know everything. And then get into fights about it in the comments. I've had people send horrible messages to me, people get really heated about stuff on [Facebook].*" In contrast to this social media experience, P2 said, "*I think that definitely would be eliminated in a forum like [SleuthTalk] because you can just have people that you know, aren't going to go like nuts.*"

Additionally, P6 shared a concern that he might lose control of the photo he uploaded to the Facebook Civil War discussion group, especially when the photo can worth a lot of money. P6 explained, "*Someone just sort of right-clicks and takes your image or does a screenshot and take your image and uses it to make a counterfeit copy or use it in their publication or use it without attribution. I know copyright rules are different. [But] you lose a measure of control [when] you post something on Facebook…*"

Besides counterfeiting, P6 also voiced concerns about theft: *"It's not always ideal to just stick [the photo] you have that's worth some money on a public forum and everyone now knows you have it.*" Likewise, P4 also shared an example where collectors have a large collection that is worth a lot of money and they do not want to let everyone know that they have it. In SleuthTalk, all of the conversation and the sources for the photo are restricted to

the people whom the project creator invites. The low barrier to entry of Facebook was seen by some as having downsides. P6 explained that he feels much better to use the SleuthTalk in CWPS because, compared to Facebook, "*you have to sort of show the commitment [to use the website].*" In other words, the specific focus of CWPS on Civil War history (versus the broad audience and generic focus of some social media platforms) might attract a more dedicated, helpful group.

## 5.5 Structured Feedback

### 5.5.1 Users voted on both individual comparisons and across candidates.

According to the log analysis, we observed that participants heavily used all of the voting features available in SleuthTalk, including the individual comparisons and identification votes (for each candidate) and the overall poll where they voted for a single winner. Across all nine projects in the study, we found, in total, 152 comparison votes and 125 identification votes. In addition, we also found that 48% of the project members voted across candidates in a poll.

### 5.5.2 Users find that SleuthTalk provides an organized and structured way to see others' opinions.

In SleuthTalk, we designed a workflow for users to follow steps to identify the unknown photo. The first step is to compare all the candidates and provide both comparison and identification votes. According to the logs, we have found that all participants have successfully finished

| Predefined Feature Name | Hair | Nose | Eye(s) | Eyebrow | Jawline | Facial Hair | Ear(s) |
|---|---|---|---|---|---|---|---|
| # of Usage | 47 | 40 | 33 | 31 | 29 | 24 | 20 |

Table 5.1: Usage of the predefined facial feature selection

this first step in their projects.

In addition, we also observed that users labeled a variety of facial features, totalling 254 selections, before they vote on the facial similarity. Moreover, 89% of the selected facial feature are from the predefined categories: hair, nose, eye(s), eyebrow, jawline, facial hair and ear(s) (see table 5.1).

Then, after all the discussion and source sharing, we expect users to start the poll and vote on the winning candidates. However, not every poll in the projects has been started. Sometimes polls were not started because there were no similar candidates. P2 explained, "*I didn't think it was really necessary. Because it seems everyone is pretty confident that they weren't [the match].*"

Overall, participants expressed the easiness of understanding the whole structure in the SleuthTalk. In the words of P5, "*it was easy to understand how it was, how it is layered and how it is structured…There's mechanisms in place to measure and there's opportunities to get the quantitative information but also have the anecdotal discussion.*"

### 5.5.3 Users think SleuthTalk is more focused and helps to filter out irrelevant information.

Many participants shared their experience of using the Facebook group to identify the unknown photo. However, their experience of using the discussion feature on Facebook was not as good as they expected it to be. P3 explained that Facebook discussions can be unfocused: "*It just seems like sometimes people miss the mark, they just kind of trail off after a while.*"

P4 stated similar reasons in that "*you might have some totally random person that just happens to be a forum member of Civil War Faces [the Facebook group]. They're just coming to look and they see it, and they might leave a [reply] in their comment, and [it] might not even be relevant to the actual discussion.*" Besides completely irrelevant comments, participants also complained about Facebook users asking to buy photos instead of providing the requested information. P5 shared his experience in this. He said, "*it's just frustrating to me [that] people will message you [if they] want to buy things from you. Sometimes it's not a great experience.*"

In contrast, participants expressed that the discussion in SleuthTalk is more focused and has higher quality. P6 credited the ability to choose one's audience: "*the discussion is helpful [in SleuthTalk] because I think you're sort of ensuring a higher quality of people that are in the discussion.*" Another participant found the structured interface helped focus discussion. P5 said that one thing "*I really enjoyed about it was [that] it sort of default filters out the irrelevant conversation.*"

## 5.6 Holistic Identification Process

### 5.6.1 Selecting facial features is helpful but insufficient on its own to show the facial similarity.

As Mohanty et al. found in Second Opinion [29], selecting specific facial features helps the user to see better while comparing two different photos. Our results confirm these earlier findings. In the words of P2, "*I can't differentiate faces very well, [but looking at the facial features] is really helpful to me, [and] if you add a box around [the selected facial feature], somehow I'm able to see it better.*" Furthermore, participants also mentioned that getting a

second opinion from others can help fill out the gap when users overlook the features. In the words of P3, "*if one [holistic] picture had facial hair and one didn't, then you can actually look up a little bit closer and see like contours of their jaw line and other different features that from far away, or when [you] compare them, you might overlook.*"

However, even though participants expressed the effectiveness of looking into the facial feature, they found it is not sufficient enough on its own to determine the facial similarity. Beyond the individual feature comparisons, one must consider the overall, holistic similarity. As P1 explained, "*I had one candidate that had his features, the nose, the lips, everything, eyebrows were very similar. But taking a step back and looking at the face, it just wasn't the same person.*"

### 5.6.2 Facial similarity is helpful but insufficient on its own to determine the final identity.

Above, we noted that Project 4 contained four poll votes on one of the shortlisted candidates, which also has six identification votes for "not sure". Even though many participants expressed that the unknown photo looks very similar to this candidate, in all the responses, none of the project creators would go ahead to finally identify the unknown photo as this winning candidate. Explaining this situation, P4 stated that for him, facial recognition only weighs 65% of the process to confirm the final identity because he believes the facial feature might change with the time goes and age changes. But if there are multiple similar photos pointing to the same person, he said, "*maybe it brings it up to 85% or 90%. The more photos that there are, the stronger there is the percentage of verification.*"

In addition to using different similar photos to prove the identity, some participants also mentioned facial similarity is not enough to make a final decision. For example, some par-

ticipants emphasized the importance of supporting documentation, such as military records. In the words of P2, this documentary "*evidence to me is the most important thing like way beyond facial matches.*"

### 5.6.3 Other users' feedback is helpful but insufficient on its own to determine the final identity.

Many participants mentioned that it is tempting to just look at what they wish to look at, and ignore anything else, during the process of identifying the unknown Civil War photos. In these situations, the perspectives provided by other users in SleuthTalk is helpful for fighting this confirmation bias. In the words of P3, "*sometimes you get those blinders on because you want it to be that person so bad. …if a couple of other people say something, it gives you a second look, and helps you.*"

However, while others' opinion are helpful for users to look in alternative ways, many participants only treated this feedback as a suggestion, and continued looking for more evidence to confirm the identity. For example, in Project 4, created by P4, even though there are four out of six votes for one candidate in the poll, P4 decided to hold off and wait for new evidence to prove it. P5 said regarding this project member feedback, "*I do take that into consideration, and I might be the only dissenting voice. I kind of want to still leave it open for interpretation. I think the more evidence you have, it very much helps the photo.*"

## 5.7 SleuthTalk vs. Facebook

### 5.7.1 Participants think that SleuthTalk serves different, but complementary, purposes from Facebook groups.

During the interviews, half of the participants self-reported of receiving the identity of their unknown photo during the Facebook task in the study. Taking advantage of the large membership of the Facebook Civil War group (over 13,000 members), many participants found it is beneficial to post there to get an idea of the identity or other information about the unknown photo when they have no idea about it. As P2 explained, "*I just want to get a broad sense if anyone recognizes, like this guy, or has a photo of this guy in their collection.*" In addition, P4 also believes that there might be many people who collect different types of Civil War photos and can contribute information about the one he posted. He said, "*<collector's name on Facebook> knows New Hampshire images, so if you have an image that has a backmark from New Hampshire, you would have a good chance running it by him, because it's very possible he might know who that guy is, just by memory, because he has been studying.*"

Beyond the Facebook group, participants also shared the times they would prefer to use SleuthTalk. Most often, they saw SleuthTalk as superior to Facebook when they already have the potential candidates in mind. In the words of P2, "*if I have three [candidates] that I think are all really like [the photo I uploaded], I'm gonna ask very direct people get their opinion on just the faces. And I don't need historical context or to reach out to wider collectors or anything, [such as the Facebook group].*" Participants saw SleuthTalk as better suited for discussion and debate around a shortlist of top candidates, but Facebook was better for crowdsourcing initial suggestions about the identity of an unknown photo. The

sites could also be used in a complementary way, where Facebook is used initially to solicit potential IDs, and then the discussion moves to SleuthTalk for more focused discussion of the candidates.

### 5.7.2 Participants think it is faster to receive a response from the Facebook group than SleuthTalk, but the request is more likely to be ignored on Facebook.

P4 shared his experience with us that he got the identity of his unknown photo in a very short time on Facebook. In the words of P4, "*I was surprised because it got identified really quickly.*" P1 also mentioned this point, and she explained the reason why she thinks it is slower to get the response in SleuthTalk. She said, "*Facebook is more well-known than [the SleuthTalk] at this time, so if they're on Facebook, that's going to be an automatic, fast response. [But for SleuthTalk], it might be a slower response.*" Since SleuthTalk was only available to study participants and the project members they invited, and the CWPS registered user community is comparable in size to the Facebook group, this concern may be partly mitigated following a public release. P1 also observed that Facebook provides a wider range of services beyond Civil War photo discussion, so users might check it more often, whereas SleuthTalk is a more narrowly focused website: "*Unless you're in the project or used to the [website], just because people might be working and they'll have to see the email, to [join the project and] respond.*"

However, contradicting to her initial thoughts, P1 expressed that she did not get many responses about the photo she posted to the Facebook group. She said, "*People responded on the Facebook page, and it seems like a lot of them. [But] I put a comment out there, I didn't always get a response, maybe the other members knew each other or had interacted*

*before.*" P1 believes that she is less likely to be ignored in SleuthTalk because she would only invite someone she knows: "*I was kind of going for a targeted audience of people that I knew had an interest in this kind of thing.*"

# Chapter 6

# Discussion

In the introduction to this thesis, we sought to address the following research questions:

- **RQ 1:** What system features allow for a software tool to address the last-mile problem of historical person identification?
- **RQ 2:** How effective are these features at allowing users to address the last-mile problem of historical person identification?

In the following subsections, we discuss the key system features that SleuthTalk provided to address the last-mile problem — i.e., shortlist creation and management, small private workspaces, and structured feedback — and their effectiveness as shown during our evaluation study.

## 6.1 Shortlist creation and management

To help reduce false positives, Mohanty et al. [30] designed CWPS to only display the Civil War photo results which satisfy the search filters and have a similarity score of 0.50 (out of 1) or higher compared to the query photo, sorted by the similarity score in a descending order. Therefore, when users visit the search results page, the first few candidates are the most similar ones based on the facial recognition algorithm. However, in a follow-up

benchmarking study, Mohanty et al. [31] showed that the highest-similarity candidates in the search result are not guaranteed to match the unknown photo. Indeed, the correct match is sometimes ranked lower, and in one illustrative example from the study, there was a photo with a match that was ranked 148th in the search results based on facial similarity.

Our study supports these earlier findings in that we find that users' shortlisted candidates may or may not be the top AI matches on the search results page. Many participants also would spend some time going through every result, and the candidates they added to the shortlist could be at the very end of the list. Therefore, the SleuthTalk evaluation showed that participants are not limited to the order and results that AI provided; instead, they are free to manage their own shortlist by just taking the facial recognition results as a suggestion.

Furthermore, after users have created a shortlist in a project, we found that some wanted to change it later. One participant added another candidate because he was too busy to go through all the search results in the original session. Another participant said that he would add a new candidate that he found outside of the CWPS to build a stronger argument. At the same time, since none of the projects is finished, every participant expressed that they would archive and add candidates in the future, after our study ended, though we do not have observational data to confirm this. Based on participants' actions and plans to identify the unknown photo in the future, SleuthTalk provides a way for users to correct the initial shortlist by allowing them to add or update candidates. While the archiving feature allowed them to remove candidates, we also found that participants were hesitant to remove candidates for several reasons and did not archive any during the study.

Our evaluation also found that four participants received a suggested identity from other users on Facebook, and three of them self-reported that their unknown photo got identified. Out of all three of them, one participant reported receiving an identity with an identical photo (i.e., replica), and two participants reported receiving an identity with a different photo of

the same soldier (i.e., facial match). To combat the confirmation bias, SleuthTalk can play an important role in this latter situation. Because two participants received a suggested identity with a different photo, thus requiring verification via second opinions, users can add all the suggestions for the identity to their shortlist in the SleuthTalk workspace and invite others to help with confirming the answers.

For the remaining two participants or other users who do not receive any potential candidates suggested by the community, facial recognition results from CWPS can also help to provide a list of the similar photos in a descending order. In the interviews, participants shared their strategies for comparing the facial similarity and checking the biographical information to narrow down search results after all the photos have been filtered by visual tags.

Therefore, as we can see that SleuthTalk is intended to help with getting a second opinion, i.e., confirming an identity, in a small pool of candidates in the shortlist, while the Facebook Civil War group is focused on coming up with an new identity from a blank slate. The two sites can complement each other in providing different types of value.

## 6.2 Benefits and trade-offs of having a small group private workspace

Misinformation is an ongoing problem in many different fields, including historical person identification. Sometimes, users can notice the mistake easily if the photo is a famous person. However, sites like CWPS focus on relatively unknown historical persons who can not be obviously identified or misidentified. In this context, many times, users are 'fooled' by a misidentified photo. To address this problem, we aimed with SleuthTalk to help to reduce the spread of misinformation during the identification process in the Civil War photo

community.

Although the Facebook Civil War group is a private group, it is very large, containing around 13,000 users, a big proportion of the overall community. In this sense, it is essentially public. Therefore, if a discussion is open to many people, such as in the Facebook Civil War group, it is more likely that other people will take the information out of context (i.e., to another website) and treat it as the identity of the unknown photo, while it needs more examination to confirm the answer. According to our interviews, even though participants are very careful not to misidentify any photo or spread misinformation, they sometimes do so unintentionally while sharing their opinions. It is hard to know whether the conversation is still ongoing or not. However, SleuthTalk helps prevent this situation by creating a private workspace with limited membership and having a clear status and steps to confirm the final identity at the end of the project.

Even though we did not see any participants reaching a conclusion about the unknown photo's identity and closing the project in SleuthTalk, this private workspace helped combat users' confirmation bias and maintain a high rate of true negatives rather than high false positives. For example, in Project 5, we showed that SleuthTalk successfully helped to change the opinion of the project creator, P3, from a positive vote to negative vote. Similarly, P4, as creator of Project 4, stopped sharing a potential misidentified match to the public as a result of the SleuthTalk project.

Furthermore, Civil War photos can have a significant monetary value. Many collectors expressed that an identified photo is usually worth much more money than an unidentified photo; the identified image can increase in value 50% or more [1]. Therefore, some users who are collectors or dealers (i.e., with financial stakes in the identification) would like to keep the final identity private to themselves. Using a public forum to determine the identity might cause the buyers to pay much more money during the auction. A related concern is

that the public becomes aware that they own the valuable photo, potentially making them a target of theft. The private feature in SleuthTalk also helps with addressing this problem. Users can take advantage of this workspace to manage potential candidates in the shortlist and collect internal and external data to determine the final identity. Nevertheless, most importantly, they can keep the final result private to themselves only.

While the private feature in SleuthTalk brings many benefits to the user, there are also trade-offs. For example, privacy restricts the number of people who can access the workspace and share their research results. Therefore, project creators may get less helpful information in a private group, while that information is vital in the Civil War photo community. On the other hand, CWPS, as a crowdsourced website that depends on its reference database, can only be more helpful if users are contributing to this community and provide much information to the public. If we make it easier for users to use this private feature to hide their photo identities from everyone, CWPS has less information to share with the community, so users will have a lower chance to match an identity through CWPS. Our hope, tentatively supported by the interviews, is that most participants will keep the project publicly only until they reach a solid conclusion. Other users who intend to keep the photo's identity permanently a secret at least have the ability to use the tools of CWPS and SleuthTalk to improve the quality of their research, and always have the option of releasing the information with the click of a button, should they have a change of mind.

## 6.3 Providing clear structure to support collaborative historical person identification

Prior work [30] has mentioned that CWPS users posted screenshots of CWPS comparisons on social media to solicit second opinions from other people in the community. Even though the photo is from CWPS, users bring the discussion to social media. Therefore, our motivation for SleuthTalk is to bring the discussion back to the CWPS, not only to centralize the discussion where other users can easily find it, but also to provide customized tools and user interfaces to better support collaborative identification.

Therefore, SleuthTalk is designed to have a structured workflow to help users perform historical person identification. It contains three steps:

1. Project members compare and vote on all the shortlisted candidates, using the facial feature selection to improve the accuracy of the facial similarity comparison.

2. Project members can discuss conflicting opinion and share other information sources outside of CWPS in the discussion section, and modify the shortlist as needed.

3. Any project member can start the poll, and everyone can vote on their selection after completing comparison voting for each candidate. After at least half of the members vote in the poll, the project creator can decide on the final identity as the winning candidate.

According to our findings, participants felt that it is easy and straightforward to understand the workflow, and all of them completed the first step in the workflow. In addition, participants also heavily used the facial feature selection, and many of them changed their

minds about the facial similarity after the first step. These results indicate that our system successfully helped users to compare the photos in detail and narrow down choices.

However, in the current version of SleuthTalk, comparison of facial features is only supported in the individual comparisons; users vote across candidates in the poll mainly based on the identification votes and external sources shared by others. In the future, to help users consider facial similarity when they vote across candidates, we would like to design a heatmap visualization to display the multi-object data [35], such as all the shortlisted candidates' photos, the same labeled features, users' opinions on similarity of the features, their comments, and their selected cropping areas. For example, in the heatmap, each column could represent one candidate's photo, and each row represents each selected facial feature's label. Then, for each of the selected facial features of a given type for the same candidate photo (e.g., all the selected features about the jawline on candidate photo 24353), we could calculate and display the 'similarity' and 'difference' percentages. If the percentage for the 'difference' opinion is higher than the 'similarity' opinion, we could use the color red to indicate that majority of the users think it is different; otherwise, we could use the color green to indicate similarity. In addition, we could use the saturation of the color to represent the percentage. The higher percentage is, the darker saturation will be displayed in the heatmap. Additionally, we could add a modal dialog box or panel next to the heatmap to display the feature selections' details, including the usernames, comments, and cropping areas, whenever user hovers on each color box in the heatmap.

At the same time, because of the heavy usage of facial feature selection and labeling, SleuthTalk creates a valuable set of human-labelled image data during each project. We could also feed this hand-curated data back into facial recognition algorithm as additional training data to improve the model (e.g., [11, 33]). For example, as users are selecting and labeling salient features, this data could inform the model about the precise boundaries and

specific labels of facial features to improve detection. Perhaps more interestingly, users are focusing on labeling the salient features (i.e., high-diagnostic features [29]) that help with differentiation between photos, and their labels are often paired across two photos; this could give the AI models deeper insights about which features are most helpful for identification (i.e., face verification). We cannot yet explore these approaches because we use a commercial face recognition service from Microsoft that does not allow for third-party model tuning, but it could be accomplished by transitioning to an open-source computer vision package such as OpenCV.

In addition, we did not see much discussion to address conflicting opinions in the workspace, or introduction of external information sources. Instead, participants gave most of their attention to the first step, facial comparison. We reflected on this result and present one observation and one potential design implication. First, many participants felt it is much faster to get responses and possibly the identity on Facebook. Because SleuthTalk is very new to the users and is not supported with a notification feature yet, people are less likely to track the workspace status all the time. Therefore, going through all the steps and finding the final identity in SleuthTalk can take longer than Facebook. However, since we only provided participants a week, which is the same amount of time as they worked on the Facebook task, participants felt that they did not get enough time to go through the remaining steps on SleuthTalk yet. All the participants expressed their excitement to keep their project open and work on it in the future.

Second, while we think that participants need more time to work on the project, we also felt that the second step of the workflow is not yet properly emphasized on the page. The current UI prioritizes the candidate voting, but discussion requires scrolling down for a while. Therefore, in future work, we would redesign the section under the shortlist and create a place to encourage users to share the other sources outside of the CWPS. At the same time,

we want to bring the comments that members provided in the first step into the discussion to encourage them to address the conflicting opinions.

## 6.4 Usage in other domains

Beyond using this collaborative tool for historical person identification, a modified version of SleuthTalk could potentially also be used in other domains. For example, in the fields of the law enforcement or national security, investigators could use a tool like SleuthTalk to narrow down candidates for a suspected criminal, witness, or missing person. Reports from crowds or other information and records from inside or outside the agency could provide broader context to complement the facial comparisons, and investigators can also manage the shortlist as more information comes in. Additionally, this such tool can provide a small, private workspace which has controlled access in and out, protecting sensitive information and preventing misidentifications from leading out, while also ensuring that all the data provided by other project members remains focused and valuable to the project creator. Furthermore, this collaborative tool can also provide an organized way to display all the voting data, including both individual comparison and across candidates poll. Therefore, everyone's thoughts are collected and people are less likely to omit anyone's opinion if he or she is the only one oppose to others.

# Chapter 7

# Conclusion

In the Civil War photo community, identifying unknown photos remains challenging because of the large pool of candidates. While AI-based face recognition can narrow down this large pool of candidates to a smaller number of possibilities, members of this community still struggle to decide among the last mile of highly similar candidates. In this work, we are contributing to addressing the last-mile problem in historical person identification by designing and presenting the collaborative tool, SleuthTalk, to provide a free and private workspace for users to interact with each other and a clear, structured workflow for users to identify the unknown historical photo. Beyond just getting a second opinion from others about the facial features [29], we designed three different components to achieve these goals: shortlisting potential candidates, private collaborative workspace, and structured feedback.

In addition, we also conducted a mixed-method evaluation with 6 participants who spent one week using the SleuthTalk and another one using a popular Facebook group to identify unknown Civil War photos. The results showed that the Facebook Civil War group serves a different purpose than SleuthTalk, and the two platforms may serve complementary functions. However, we successfully helped address some of its issues, such as the unfocused and disorganized discussion during the identification process. We also contribute to preventing the spread of the misinformation in the Civil War faces community by designing the workspace to be temporarily private.

We believe that our work will help future researchers better understand the challenges and

provide design implications when building a collaborative tool to support historical person identification.

# Bibliography


[1] How artificial intelligence is helping identify thousands of unknown civil war soldiers. URL https://time.com/5749059/civil-war-photos/.

[2] Ancestry - genealogy, family trees & family history records. URL https://www.ancestry.com.

[3] Dulles cbp's new biometric verification technology catches third impostor in 40 days. URL https://www.cbp.gov/newsroom/national-media-release/dulles-cbp-s-new-biometric-verification-technology-catches-third.

[4] What is the face recognition setting on facebook and how does it work? | facebook help center, . URL https://www.facebook.com/help/122175507864081.

[5] Nijeer parks was arrested due to a false facial recognition match - cnn, . URL https://www.cnn.com/2021/04/29/tech/nijeer-parks-facial-recognition-police-arrest/index.html.

[6] Find a grave - millions of cemetery records. URL https://www.findagrave.com.

[7] Photography as historical documentation 2018, Dec 2018. URL https://worldhistoryarchive.wordpress.com/2018/12/18/photography-as-historical-documentation.

[8] Facial recognition technology leads to arrest in las vegas killing, Apr 2020. URL https://www.reviewjournal.com/crime/homicides/facial-recognition-technology-leads-to-arrest-in-las-vegas-killing-2018249.

[9] Natanael Arndt, Kurt Junghanns, Roy Meissner, Philipp Frischmuth, Norman Radtke, Marvin Frommhold, and Michael Martin. *Structured Feedback: A Distributed Protocol for Feedback and Patches on the Web of Data*. Apr 2016.

[10] Duen Horng Chau, Aniket Kittur, Jason I. Hong, and Christos Faloutsos. Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 167–176. Association for Computing Machinery, May 2011. ISBN 978-1-4503-0228-9. doi: 10.1145/1978942.1978967. URL https://doi.org/10.1145/1978942.1978967.

[11] 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*, page 600–611. ACM, Feb 2015. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675214. URL https://dl.acm.org/doi/10.1145/2675133.2675214.

[12] Gregorio Convertino, Dorrit Billman, Peter Pirolli, and Jeff Shrager. Collaborative intelligence analysis with cache: Bias reduction and information coverage. Jan 2005.

[13] Maia B. Cook and Harvey S. Smallman. Visual evidence landscapes: Reducing bias in collaborative intelligence analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 51(4):303–307, Oct 2007. ISSN 2169-5067. doi: 10.1177/154193120705100433. URL https://doi.org/10.1177/154193120705100433.

[14] Anthony DeBartolo. Appraisers can tell when old photo really may be golden. URL https://www.chicagotribune.com/news/ct-xpm-1985-12-27-8503300067-story.html.

[15] Zhaoan Dong, Jiaheng Lu, Tok Wang Ling, Ju Fan, and Yueguo Chen. Using hybrid algorithmic-crowdsourcing methods for academic knowledge acquisition. *Cluster Computing*, 20(4):3629–3641, Dec 2017. ISSN 1573-7543. doi: 10.1007/s10586-017-1089-8. URL https://doi.org/10.1007/s10586-017-1089-8.

[16] Shelly Farnham, Harry R. Chesley, Debbie E. McGhee, Reena Kawal, and Jennifer Landau. Structured online interactions: improving the decision-making of small discussion groups. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work - CSCW '00*, page 299–308. ACM Press, 2000. ISBN 978-1-58113-222-9. doi: 10.1145/358916.359001. URL http://portal.acm.org/citation.cfm?doid=358916.359001.

[17] H. Galiyawala, K. Shah, V. Gajjar, and M. S. Raval. Person retrieval in surveillance video using height, color and gender. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, page 1–6, Nov 2018. doi: 10.1109/AVSS.2018.8639145.

[18] Christine Geeng, Savanna Yee, and Franziska Roesner. Fake news on facebook and twitter: Investigating how people (don't) investigate. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–14. ACM, Apr 2020. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376784. URL https://dl.acm.org/doi/10.1145/3313831.3376784.

[19] Anhong Guo, Anuraag Jain, Shomiron Ghose, Gierad Laput, Chris Harrison, and Jeffrey P. Bigham. Crowd-ai camera sensing in the real world. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):111:1–111:20, Sep 2018. doi: 10.1145/3264921. URL https://doi.org/10.1145/3264921.

[20] E. Hossain and G. Chetty. Person identification in surveillance video using gait biometric

cues. In *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, page 1877–1881, May 2012. doi: 10.1109/FSKD.2012.6234146.

[21] Binxuan Huang and Kathleen M. Carley. Disinformation and misinformation on twitter during the novel coronavirus outbreak. *arXiv:2006.04278 [cs]*, Jun 2020. URL http://arxiv.org/abs/2006.04278. arXiv: 2006.04278.

[22] Vi Nguyen • Published January 15, 2021 • Updated on January 15, and 2021 at 5:14 Pm. 'needs to stop:' internet users misidentify retired chicago firefighter as riot suspect. URL https://www.nbcchicago.com/news/local/needs-to-stop-internet-users-misidentify-retired-chicago-firefighter-as-riot-suspec 2417255/.

[23] Aniket Kittur, Duen Horng Chau, Christos Faloutsos, and Jason I Hong. Supporting ad hoc sensemaking: Integrating cognitive, hci, and data mining approaches. page 4.

[24] Aniket Kittur, Andrew M. Peters, Abdigani Diriye, and Michael Bove. Standing on the schemas of giants: socially augmented information foraging. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, page 999–1010. ACM, Feb 2014. ISBN 978-1-4503-2540-0. doi: 10.1145/2531602.2531644. URL https://dl.acm.org/doi/10.1145/2531602.2531644.

[25] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. Supporting reflective public thought with considerit. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, page 265. ACM Press, 2012. ISBN 978-1-4503-1086-4. doi: 10.1145/2145204.2145249. URL http://dl.acm.org/citation.cfm?doid=2145204.2145249.

[26] Sung-Chul Lee, Jaeyoon Song, Eun-Young Ko, Seongho Park, Jihee Kim, and Juho Kim. Solutionchat: Real-time moderator support for chat-based structured discussion.

In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, page 1–12. ACM, Apr 2020. ISBN 978-1-4503-6708-0. doi: 10.1145/3313831.3376609. URL https://dl.acm.org/doi/10.1145/3313831.3376609.

[27] Weichen Liu, Sijia Xiao, Jacob T. Browne, Ming Yang, and Steven P. Dow. Consensus: Supporting multi-criteria group decisions by visualizing points of disagreement. *ACM Transactions on Social Computing*, 1(1):1–26, Feb 2018. ISSN 2469-7818, 2469-7826. doi: 10.1145/3159649. URL https://dl.acm.org/doi/10.1145/3159649.

[28] Kurt Luther, Jari-Lee Tolentino, Wei Wu, Amy Pavel, Brian P. Bailey, Maneesh Agrawala, Björn Hartmann, and Steven P. Dow. Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, page 473–485. ACM, Feb 2015. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675283. URL https://dl.acm.org/doi/10.1145/2675133.2675283.

[29] Vikram Mohanty, Kareem Abdol-Hamid, Courtney Ebersohl, and Kurt Luther. Second opinion: Supporting last-mile person identification with crowdsourcing and face recognition. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7(11):86–96, Oct 2019. URL https://ojs.aaai.org/index.php/HCOMP/article/view/5272.

[30] 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*, page 547–557. ACM, Mar 2019. ISBN 978-1-4503-6272-6. doi: 10.1145/3301275.3302301. URL https://dl.acm.org/doi/10.1145/3301275.3302301.

[31] Vikram Mohanty, David Thames, Sneha Mehta, and Kurt Luther. Photo sleuth: Identi-

fying historical portraits with face recognition and crowdsourced human expertise. *ACM Transactions on Interactive Intelligent Systems*, 10(4):1–36, Dec 2020. ISSN 2160-6455, 2160-6463. doi: 10.1145/3365842. URL https://dl.acm.org/doi/10.1145/3365842.

[32] Davide Parmigiani. Teachers and decision-making processes: An italian exploratory study on individual and collaborative decisions. *Canadian Journal of Education / Revue canadienne de l'éducation*, 35(1):171–186, 2012. ISSN 0380-2361. URL https://www.jstor.org/stable/canajeducrevucan.35.1.171.

[33] G. Patterson, Grant Van Horn, Serge J. Belongie, P. Perona, and James Hays. Tropel: Crowdsourcing detectors with minimal training. In *HCOMP*, 2015.

[34] Shih-Ming Pi, Chen-Huei Chou, and Hsiu-Li Liao. A study of facebook groups members' knowledge sharing. *Computers in Human Behavior*, 29(5):1971–1979, Sep 2013. ISSN 0747-5632. doi: 10.1016/j.chb.2013.04.019. URL https://www.sciencedirect.com/science/article/pii/S0747563213001222.

[35] Andy Pryke, Sanaz Mostaghim, and Alireza Nazemi. *Heatmap Visualization of Population Based Multi Objective Algorithms*. Jan 2006. ISBN 978-3-540-70927-5. doi: 10.1007/978-3-540-70928-2_29.

[36] Jeffrey Rzeszotarski and Aniket Kittur. Crowdscape: interactively visualizing user behavior and output. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, page 55–62. Association for Computing Machinery, Oct 2012. ISBN 978-1-4503-1580-7. doi: 10.1145/2380116.2380125. URL https://doi.org/10.1145/2380116.2380125.

[37] Christina Schwind and Jürgen Buder. Reducing confirmation bias and evaluation bias: When are preference-inconsistent recommendations effective – and when not? *Computers in Human Behavior*, 28(6):2280–2290, Nov 2012. ISSN 0747-5632. doi: 10.

1016/j.chb.2012.06.035. URL https://www.sciencedirect.com/science/article/pii/S0747563212001963.

[38] Constanza Villarroel, Mark Felton, and Merce Garcia-Mila. Arguing against confirmation bias: The effect of argumentative discourse goals on the use of disconfirming evidence in written argument. *International Journal of Educational Research*, 79: 167–179, Jan 2016. ISSN 0883-0355. doi: 10.1016/j.ijer.2016.06.009. URL https://www.sciencedirect.com/science/article/pii/S0883035516300453.

[39] Heather L. Willever-Farr and Andrea Forte. Family matters: control and conflict in online family history production. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, page 475–486. ACM, Feb 2014. ISBN 978-1-4503-2540-0. doi: 10.1145/2531602.2531737. URL https://dl.acm.org/doi/10.1145/2531602.2531737.

[40] Anbang Xu and Brian Bailey. What do you think?: a case study of benefit, expectation, and interaction in a large online critique community. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work - CSCW '12*, page 295. ACM Press, 2012. ISBN 978-1-4503-1086-4. doi: 10.1145/2145204.2145252. URL http://dl.acm.org/citation.cfm?doid=2145204.2145252.

[41] Anbang Xu, Shih-Wen Huang, and Brian Bailey. Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, page 1433–1444. ACM, Feb 2014. ISBN 978-1-4503-2540-0. doi: 10.1145/2531602.2531604. URL https://dl.acm.org/doi/10.1145/2531602.2531604.

[42] Daniel Zhang, Yang Zhang, Qi Li, Thomas Plummer, and Dong Wang. Crowdlearn: A crowd-ai hybrid system for deep learning-based damage assessment applications. In

*2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, page 1221–1232, Jul 2019. doi: 10.1109/ICDCS.2019.00123.

[43] Roshanak Zilouchian Moghaddam, Zane Nicholson, and Brian P. Bailey. Procid: Bridging consensus building theory with the practice of distributed design discussions. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, page 686–699. ACM, Feb 2015. ISBN 978-1-4503-2922-4. doi: 10.1145/2675133.2675272. URL https://dl.acm.org/doi/10.1145/2675133.2675272.

# Appendices

# Appendix A

# Interview Questions

1. Photo
   (a) Where is it from?
   (b) Have you tried to identify before?
2. Overall Experience
   (a) How was your experience identifying photo on SleuthTalk?
   (b) How was your experience identifying photo on Facebook?
   (c) What did you expect? Did it meet your expectations?
   (d) What did you like and dislike about these two platforms?
3. Privacy
   (a) How important is privacy for your Civil War photo identifications?
   (b) In terms of the privacy setting, which one would you prefer to use? SleuthTalk or Facebook? Can you tell us more about that?
4. Discussion
   (a) How was your experience when you were using the discussion on both SleuthTalk and Facebook?

(b) How was all the information organized in both platforms for you?

5. Team Members

   (a) How did you decide who to invite?

   (b) What information did you provide when you invite others?

   (c) How did you decide which project to join?

6. Leader vs. Member

   (a) To Leader: What was your strategies to lead the group to identify the unknown photo?

   (b) To Leader: How involve do you think you were in the group?

   (c) To Member: What was your expectations to project leaders?

   (d) To Member: How involve do you think you were in the group?

   (e) What do you think about your responsibility to keep moving the project process going forward? (e.g. start or reset the poll, etc.)

   (f) What is your preference on being a leader or member in the group?

7. Shortlisted Candidates

   (a) What factor did you have to pick and manage your shortlisted candidates?

8. Comparison and Voting

   (a) Did you aware of starting the comparison vote?

   (b) How did the process of the comparison vote help you to compare with two different photos?

   (c) How did other people's opinion influence you to vote?

(d) Did you aware of starting the poll?

(e) How did the poll help you in the process of identifying the unknown photo?

(f) How did you make the final decision to identify your unknown person to a candidate? (e.g. Everyone's vote? Your initial thought? etc.)

9. In what situation will you choose to use SleuthTalk or Facebook group in the feature?

10. Do you have anything else you would like to let us know about your experience?

11. Do you have any questions for us?