Aviation History Research Network White Paper: Facial Recognition as a Tool for Archival Research

Johnathan Clayborn, MS – Director of Invader Historical Foundation

ABSTRACT: *This paper discusses the feasibility of using facial recognition technology as a tool in historical research, including discussions on efficacy, scale, benefits, and limitations of use. It is the opinion of this author that this technology in its present iteration may provide some useful, tangible results for historical researchers, however, the results are not infallible and should always be closely checked by a skilled researcher. Caution should be exercised, and every match should be verified before being accepted.*

BACKGROUND

Much of the scholarly research on this subject centers around either concepts of uniqueness of human physiology of facial features, or around applications of this technology for law- enforcement and security purposes. To our knowledge, no other small museums or small historical research groups have published findings in this field, which makes the lessons that we have learned somewhat unique in perspective.

This paper will discuss some of the unique benefits, challenges, and processes that have been pioneered by Johnathan Clayborn in his attempts to use these tools for historical research.

Many people are familiar with the concept of facial recognition technology thanks to its prolific appearance in movies and television shows. It is most often depicted in the crime and thriller genres where a photo can be scanned, the face can be mapped, and then checked against a master database to determine who is in the picture. This makes for exciting television, but the real-life value and limitations of the technology are far more complex.

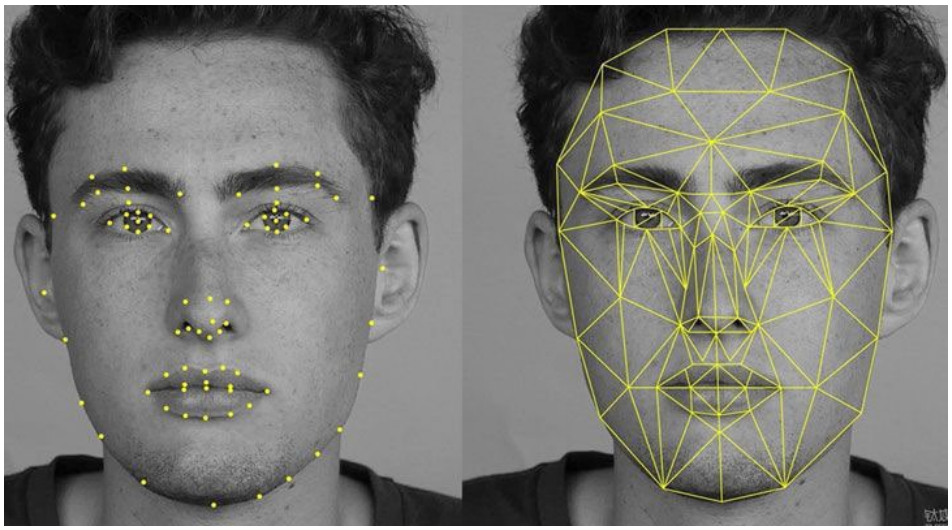
Facial recognition software uses computer algorithms to examine and evaluate a person's face for specific, unique details. These details can include things such as the length of the nose, the distance between the pupils, the width of the eyes, the distance and angle between the eyes and the corners of the mouth, the distance and angle between the corner of the mouth and nose, etc. These unique points are mapped and turned into a series of lines and points with numeric values called a template. The template is the only thing used by the software to make an identification (Electronical Frontier Foundation (EFF), 2022), (Prosecutors Center for Excellence (PCE), 2019).

Different facial recognition software produces different types of outcomes. Some facial recognition programs will tell you if the photo's template is a positive match to the known sample on file. Other systems will use a ranked percentage result and may display a list of several possible matches with the most likely match appearing first (EFF, 2022).

The following image depicts how modern facial recognition features are mapped into a template. Each point on the face is marked based on a pre-determined set of criteria, and then the lines between those points are drawn and measured, both in length and angle. The higher the resolution of the source image, the more dots can be accurately mapped, thus giving a more accurate result.

Aviation History Research Network White Paper: Facial Recognition as a Tool for Archival Research

Johnathan Clayborn, MS – Director of Invader Historical Foundation



ERRORS

There are two primary types of errors that occur with the use of Facial Recognition Technology. The first of these errors is a “false negative”. This occurs when the facial recognition software fails to match a template to a known sample. (EFF, 2022), (PCE, 2019). In this case, you have a known-good photo of the subject in your database that is properly tagged and coded. The new sample image also contains the same subject, but the facial recognition software fails to produce a match. This is most often the result of system limitations (discussed in detail later in this paper).

The second error is a “false positive”. This is when the facial recognition technology asserts that two sample images contain the same person, but they are, in fact, different people. (EFF, 2022), (PCE, 2019). In this case an image of an unknown subject can be mistakenly linked to a completely different person because they happen to share enough similarities in their facial features. This error is also due to the system limitations, discussed later.

LIMITATIONS

One of the limitations of facial recognition technology is that the recognition data can be prone to error, which can create false positives. Facial recognition software is particularly bad at recognizing persons of color and other ethnic minorities, women, and young people, often misidentifying or failing to identify them. For the sake of this paper, any photos that a researcher may be trying to identify featuring people of these groups may be problematic (EFF, 2022). However, some elements of these limitations can be addressed through more advanced AI software, machine learning, and software training (CSIS, 2022).

Facial recognition systems often struggle to make an identification when the photos are less than optimal, which, unfortunately, is often the case for those doing historical research. Some of the common factors which can hinder photo analysis include poor lighting, low image resolution, or poor angle of the subject of the photo (EFF, 2022).

Another limitation of some systems is that the system will not recognize certain faces. For example, in the photo depicted here, Google Photos only recognizes a single face in this image, despite there being 8 people in the image. This is due to combination of factors of low resolution, poor lighting, and sub-optimal angles. This image is representative of many of the types of images that will be encountered by researchers. Researchers are not able to manually select or enter a face, the rest of these subjects are invisible to the algorithms.

Pictured here is another example of a person who is not detected by the facial recognition algorithm used by Google Photos. Of the 7 people pictured, 6 have been identified as faces, but the other man is invisible to the system. This is, again, due to the combination of low resolution, poor lighting, and the fact that the face is partially obfuscated by the hat brim, concealing several of the features that the system will use to create the facial template.

This photo is one that highlights the limitations of facial recognition technology when used in historical research applications and presents a special kind of frustration to the researcher. To a human researcher, the face is clearly visible, and in this specific example, the subject is known to the author by name, unit, and location. However, due to the low quality of the photo, and the poor lighting on the face itself the Google Photo algorithm does not recognize at face at all. Thus, despite knowing the subject's information, it is impossible to tag this photo and attribute the face properly. To add those details to the face, a researcher must be able to select the face first, which in this case, cannot be done.

BEST PRACTICES

System Evaluation: When evaluating a facial recognition system for use, it is strongly recommended to do a series of test runs on the platform. Understand that no system is perfect and that every system will generate some false positives and false negatives. At a minimum, provide the system with two or three dozen known, decent quality photos of identified subjects. Portrait shots are excellent for this benchmark. Then, feed the system a random sample of 2,000 – 3,000 images. This sample batch should include other known photos of the individuals in your known-good sample database. Ideally, you should know how many photos of Subject A you gave the system, and you can measure how many accurate results it gave you for Subject A, and then the same measure can be counted for Subject B, and so on. Once all the known subjects have been counted and measured, a score can be assessed.

It is also important, to an extent, to understand which type of errors are more permissible for your use case. For example, would you rather have the system show you false positives more often, potentially linking unrelated people, or would you rather have the system generate false negatives and not match any photos? Take this into consideration when assigning your evaluation score.

Human Oversight: Although facial recognition algorithms are continually advancing and improving, facial recognition technology should be used to aid humans in their decision-making capacity rather than replace it. Human oversight with the use, settings, and results of facial recognition technology helps to mitigate the risk of errors by double checking (CSIS, 2022).

UNIQUE CHALLENGES FOR HISTORICAL RESEARCH APPLICATION

Known-Good Source Images: One problem that is especially troubling in the field of historical research is the potential lack of known-good images. Many times, photos are presented to us in a collection of random photos, usually gathered from the estates of people who have passed away, and there is often no accompanying information regarding who is in the photograph. Or, alternately, the photo may bear a name with it, but due to the age of the photo and the camera limitations, the photo may be of low quality, making it difficult to work with.

One of the best-practice approaches in this context is to attempt to find family members of the people that you are attempting to identify. If anyone on earth has a higher-quality image of the person(s) of interest, it would be their family. If they are willing to scan a high-res copy for you, this will greatly improve your chances of finding successful matches.

Missing Context: As mentioned above, many times photos come to the researchers only after the photographer or owner of the photograph has died. And in many cases, there is little or no contextual information about the photograph recorded, so the details are a mystery.

Forensically, the photo can be examined for contextual clues; are there unit markings on the planes? Are the people wearing unit insignia? Is there a specific building or landmark which can give away it's location? Are there other known people in the photograph? Sometimes, by having the answer to one or more of these questions the rest of the answers can be determined, or at the very least estimated within a narrow range.

The best possible resolution for this context problem is to have a database that is as large as possible. However, it should be noted that some facial recognition software, especially free services, do have a limit on how many faces it will track. To this end, the number of supported faces may be a key factor in deciding which facial recognition technology to employ. If you are conducting research on one specific unit, then a facial cap of 1,000 or 2,000 faces may be entirely sufficient. However, if you are conducting broad research that encompasses multiple units, time periods, and countries, then you would need a platform that supports tens of thousands of faces.

Scattered Records: This problem is not necessarily unique to historical research, but it is a situation that historical researchers need to be acutely aware of. Due to the compartmentalized nature in which historical records are kept, there is no central database of faces. Even within the realm of photographic collections, there do not exist any single-point collections which are all-encompassing for a given topic. Photos may be spread out through multiple document types including personnel files, accident files, unit history files, news images, etc. Due to the ad-hoc nature by which photos are often acquired and added, a negative result today does not always mean that the result will be negative forever.

For example, if a researcher runs a group photo through a facial recognition system and only 3 people are identified today, that does not mean that all the people cannot be identified in future. Hypothetically, 6 months from now that same researcher may add 200 more personnel file images to the platform and then several other people in that photo may become known. To that end, it is crucially important that whatever platform you select continually runs the searches on recurring basis in the background, in this way it will find new matches over time. If the platform only executes a particular search once, then the researcher must manually repeat that search process, which can become unwieldly as more and more records are added.

CASE STUDIES

With the specific limitations and challenges presented in this paper, the reader may be forgiven in thinking that attempting to use facial recognition technology in historical research is a futile endeavor. However, this is not the case. In the following pages several real-world case studies will be presented from the Invader Historical Foundation's attempts to use facial recognition. These case studies will provide a clearer example of the benefits and possible pitfalls of this technology.

For the case studies used in this paper, all the subjects come from the 13th Bombardment Squadron historical archives, from the Korean War, in particular. This photo set is comprised of 41 separate collections of photos taken by 37 different photographers spanning the years 1946 to 1955. All total, the combined photo set contains 1,993 images. The total photographic database contains 16,193 images.

The facial recognition analysis results for all these case studies were performed by Google Photo's built-in facial recognition system, which is a free service. No other tools were used in these examples.

Col. William G. Moore:

The analysis of Col. William G. Moore's results starts with this single photo. The photo provides his name and rank, and provides a reasonably good frontal portrait shot, with adequate lighting and resolution considering the age of the photo. This photo was provided as part of the Robert Fortney Collection (1952-1953), but no additional details were present. The file name simply reads "Moore, William G. Col.", which matches what is determined from the photo. Whenever possible, photos like this should be used as a base photo for facial recognition.

Counting the photo used as the base image, the Google results returned 34 matches for Col. Moore.

This match also comes from the Fortney Collection; however, the file name is "Group & Squadron Staff". There are no accompanying notes to indicate who any of these people are, but Col. Moore is clearly visible in the front, center of this image. In this case the facial recognition software correctly identified him and gave us another known photo of Col. Moore that we might not have known otherwise.

This photo, also from the Fortney Collection, is simply labeled "Group Staff at Col. Moore's quarters". There is no additional context or names provided. However, once again, the Google's facial recognition algorithm was successful in identifying Col. Moore, despite him facing sideways relative to the camera. As with the previous photo, this is another photo that was not clear where Col. Moore was, but he has now been clearly identified.

This photo comes from the Maturkanich Collection (1952). The photo is labelled "Gunner Gregory & Driver", and it's clear that the two men in the center of the image are the intended subjects. But there, on the far-right edge of the image, is Col. Moore, who just happened to be walking by when the photo was taken. In this case Maturkanich did not tag Col. Moore in the photo's details. But now we have another positive photo, and this time from an entirely different photographer.

This photograph comes from the Ray Anastos Collection (1952). The photo is labelled "Johnnie Grubbs _ _ A. Landrum." In this case, Anastos did not know the names of the two men in the center of the photo. The third man is turned away from the camera, but the second man can be identified as Col. Moore. This adds yet another positively identified photo of the Colonel, this time from a third photographer, that would not have been known without the aid of Facial recognition.

Unfortunately, this represents the last of the positive ID photos for Col. Moore. The rest of the images were false positives. Rather than display the entire image, a collection of the misidentified faces are shown here.

As can be seen, several of the men from the false positives do share broadly similar facial features – a pointed chin, a long, thin nose, high foreheads, and similar face shapes. But these men are obvious to a human observer as not being the same person.

The benefit of the system, and something that any facial recognition system should permit, is the ability to remove these faces from Col. Moore's results and unlink them. By doing that, these faces are returned to the results pool, or they can be manually created as a new person.

Part of these mixed results are due to Google's limitations as a software, and the rest are due to the conditions within the photo themselves. With regards to the Google technology, this facial recognition system was never intended to be used by researchers who are searching through thousands of photos. Google created the facial recognition algorithm with personal end-users in mind. In reading through Google's help forum about this topic it seems that some users cannot have more than 256 different faces in the system at once, while other users can maintain as many as 1,000, but no users reported being able to support more faces than that. This means that since Google's algorithms are limited to a specific number of matches, the system is trying to shoe-horn everyone it encounters into that fixed

number of faces, which is resulting in the high number of false-positives. Shopping around for a commercial system that supports a larger number of faces will be one way to prevent this issue.

Limiting the photos to only using Google, alternate strategies may include uploading smaller batches of photos. Although the subject of this case study was one person from the Korean War, the algorithm tried to pull in results from World War II, as well as results from the 17th Bomb Wing, neither of which were correct in any way. By limiting the facial searches to one smaller group at a time it can be possible to reduce the number of false positives because those photos are removed as possible matches. However, this method may miss results where a person was, in fact, in multiple units or locations.

Maj. Estes B. Sherril :

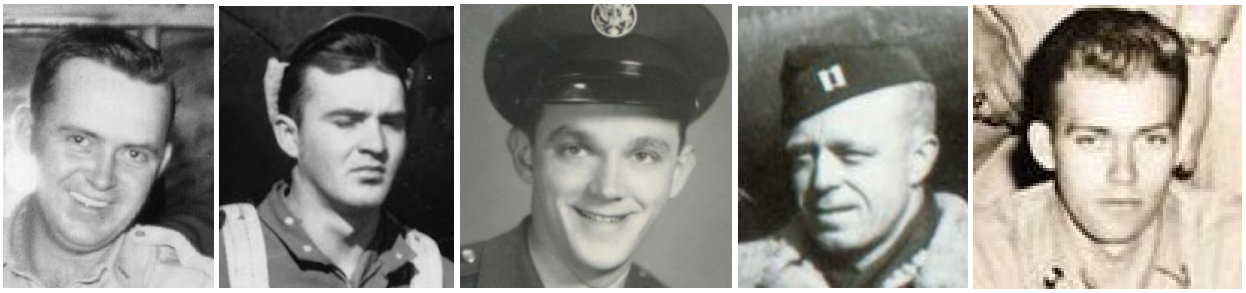
As with Col. Moore, there is a relatively clear photo of Maj. Sherrill with his face in full frontal view. The photo comes from the Fortney Collection (1952-1953) and is clearly labelled as “Sherrill, Estes B.” So here we have a good base photo to work from. (The photo has been cropped for clarity and space).

This second photo also comes from the Fortney Collection. This photo was presented as part of the Col. Moore case study (who is seated center, bottom row). The photo caption reads “Group & Squadron staff” and the facial recognition system identified Maj. Sherrill in the bottom right corner of the image. This was not a known detail without this technology.

This image (cropped) comes from the Maturkanich Collection (1952) and is merely labelled "Unknown Major #2), however, this is clearly Maj. Estes B. Sherrill. This is a third photo that would not have been known without this technology.

This image (cropped) also comes from the Maturkanich Collection (1952) and is labelled simply, "Unknown Major". Facial recognition has once again correctly identified Maj. Sherrill. This adds a 4th known photo of the major to our collection.

This is where our list of actual positive results ends. As in the case of Col. Moore, the remaining 5 results for Maj. Sherrill are false positives.



This example, as with the previous case, illustrates the importance of having a human researcher double check the facial recognition results. While some of these false positives do bear a passing resemblance to Maj. Sherrill, some of these faces are not even close. This reiterates the limitations of the system. Had a more complex, commercial grade system been employed for this test, the results would very likely be different. But since Google's algorithm is limited in the number of total faces, it tries to put every face it encounters into one of the existing profiles.

As before, many of the same techniques can be employed to combat these results; use a better system, if possible. Or, if not possible due to budgetary or technical constraints, use smaller photo sets within Google and rotate them out on a regular basis to keep the number of false positives low.

2Lt Robert J. Ramsey:

This photo is from the Ray Anastos Collection (1952) and provides a reasonably good base image for use. The photo is labelled "Ramsey, Robert J. MIA". His rank can be deduced from the pin on his cover.

This photo is from the Fortney Collection (1952-1953) and Google's facial recognition algorithm has identified the man on the far-right edge as 2Lt Ramsey. This is further supported by the presence of the same ring on the subject's left hand. In this case, the Fortney photo is labelled "Marine, Wells, Ruhlin, Murray, Hambrick, Marine, Adamack, Routh, Ramsey", so we can confirm that this is correct. But even in the event that all of the photo details were lost or missing, a positive ID of Ramsey would have been made.

Of the results present, the other 4 were false positives. In this case, the software did not identify any previously unknown photos of the subject, but it did help to confirm the IDs presented in the captions of the photos. As before, the false positives are displayed below.

These men bear some broadly similar facial features to Ramsey, but as in the other cases, to a human researcher it is evident that they are different individuals. The same mitigation techniques previously discussed are useful here as well.

Capt. Ed Hobday:

This image provides the basis for Capt. Hobday's template. It is from the Hobday Collection (1952) and the photo is labeled "Ed Hobday & Bill Holcom". Capt Hobday is on the left and the photo has remarkably good lighting and resolution.

This photo (cropped) comes from the Hobday Collection (1952) and is labelled "Ed Hobday & Russ Buker". The facial recognition algorithm correctly identified Hobday as the man on the left.

This photo (cropped) comes from the Hobday Collection (1952) and is labelled "Ed Hobday & Gunner Gene Niemi". The facial recognition algorithm once again correctly identified Capt. Hobday as the man on the left.

This photo is also from the Hobday Collection (1952) and is labeled "debriefing – Hobday, Stockton, & Buker." The Facial recognition algorithm has identified the man on the left as Capt. Hobday, which confirms the caption of the photo.

This image (cropped) also comes from the Hobday Collection (1952). It is labelled "Hobday, Stockton & Hall". The facial recognition algorithm has recognized Hobday as the man on the right. This is a somewhat surprising result as the top of his head is missing, but enough of his face is still visible for an ID.

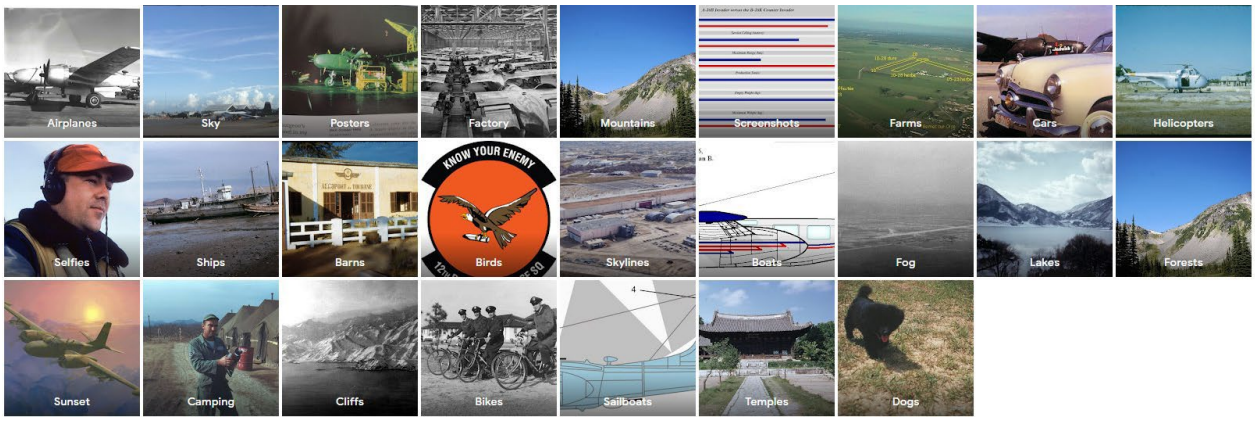
The system made a number of other positive matches for Capt. Hobday, shown here.

The system also made two false positive matches, shown below.

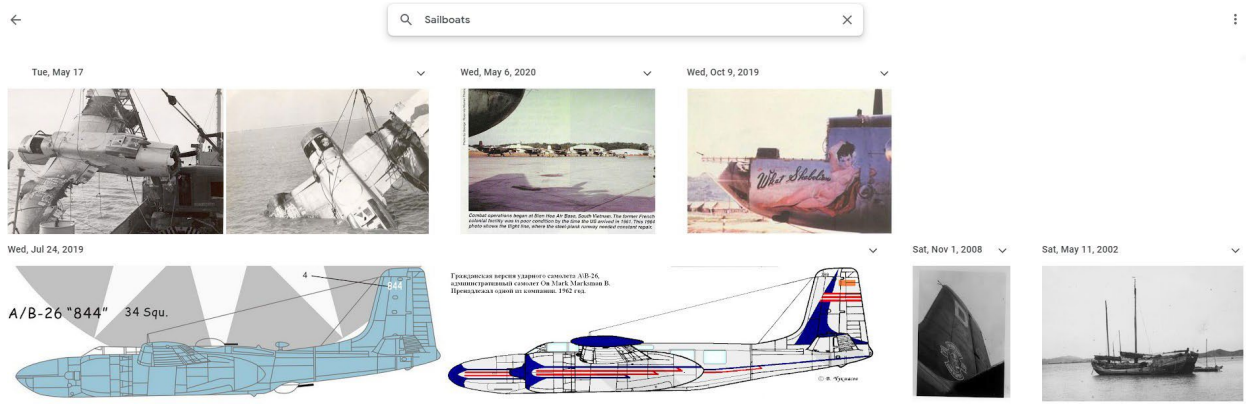
It is currently unclear exactly what combination of factors led to this lower-than-normal number of false positive errors.

Unexpected Results:

One thing that was a surprise in using the Google Photos program is that it not only attempted to group faces, but the software also attempted to create groups of similar images based on topic of the photos themselves, albeit sometimes with hilarious and unexpected results. The following list is a series of AI-created photo groups that Google created by itself from the available photos without user input.



Some of these categories are useful, and some are ones that may not have been initially considered. However, some of them are rather comical. “Camping”, for example, depicts a series of large tents that served as enlisted men’s quarters during WWII and Korea. Technically speaking, it could be regarding as long-term camping. But then there are the “sailboats”…



As can be seen, the algorithm here also returns false positives.

CONCLUSION

As with any tool in the researcher’s arsenal, facial recognition technology can be employed to some success and may assist researchers in positively identifying unknown individuals. However, the results of such work should never be blindly accepted and always verified by a human researcher. A higher degree of success can be found by utilizing best practices and by understanding the specific limitations of the platform that is being used.

SOURCES:

Center for Strategic and International Studies (CSIS) - https://www.csis.org/analysis/how-does-facial-recognition-work

Electronic Frontier Foundation (EFF) - https://www.eff.org/pages/face-recognition

Prosecutors Center for Excellence (PCE) - https://pceinc.org/wp-content/uploads/2019/11/20190528-Facial-Recognition-Article-3.pdf

FURTHER READING:

Facial Recognition Software Might Have a Racial Bias Problem –

https://apexart.org/images/breiner/articles/FacialRecognitionSoftwareMight.pdf

Batch metadata assignment to archival photograph collections using facial recognition software –

https://journal.code4lib.org/articles/8486?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+c4lj+

AI for Archives: Using Facial Recognition to Enhance Metadata –

https://research.lyrasis.org/items/b4344a34-ae78-4bbf-9549-df540a7f9aae

Bounding an archiving: assessing the relative completeness of the Jacques Toussele archive using pattern-matching and face-recognition

https://www.tandfonline.com/doi/full/10.1080/1472586X.2021.1991238

Facial Photo Recognition Using Deep Learning in Archival Record Management System

https://link.springer.com/chapter/10.1007/978-981-13-9341-9_64