Just Facial Recognition for Forensic Investigations

Introduction [00:00:01] RTI International's Justice Practice Area presents Just Science.

Introduction [00:00:10] Welcome to Just Science, a podcast for justice professionals and anyone interested in learning more about forensic science, innovative technology, current research and actionable strategies to improve the criminal justice system. On episode six of our case study season. Just Science sat down with Dr. Richard Vorder Bruegge, senior physical scientist at the Federal Bureau of Investigation, to discuss how facial recognition technology was used to help identify individuals in photographs and two landmark cases. Many individuals encounter facial recognition technology in their everyday lives, like when they unlock their phones. The same technology can significantly speed up and improve forensic investigations. Listen, along is Dr. Vorder Bruegge describes how pattern based matching is used to compare faces and images. How facial recognition technology was used to assist in the identification of unknown victims in the case against serial killer Israel Keyes and the quest to determine the identities of the Marines in the famous World War two Iwo Jima flag raising photograph. This episode is funded by the National Institute of Justice's Forensic Technology Center of Excellence. Some content in this podcast may be considered sensitive and may evoke emotional responses or may not be appropriate for younger audiences. Here's your host, Jaclynn Mckay.

Jaclynn McKay [00:01:48] Hello and welcome to Just Science. I'm your host, Jaclynn Mckay, with the Forensic Technology Center of Excellence, a program of the National Institute of Justice. On today's episode, we will discuss facial recognition technology in two case studies on how it was used to assist investigations. Here to guide us in this discussion is senior physical scientist Dr. Richard Vorder Bruegge. Welcome, Richard. Thanks for talking with us today.

Dr. Richard Vorder Bruegge [00:01:46] Thank you. Jaclynn It's a pleasure to be here.

Jaclynn McKay [00:01:48] So you currently work in the operational technology division at the FBI laboratory. Can you tell our audience a little bit about your role and your background?

Dr. Richard Vorder Bruegge [00:01:57] Absolutely. I have a bachelor's degree in engineering from Brown University. I subsequently earned a master's and a Ph.D. in geological sciences from Brown for my work in morphological and geophysical analyzes of mountain belts on the planet Venus. Comparing those processes there with those on Earth. After spending four years as a contractor for NASA and the Department of Defense, where I helped plan and actually fly robotic spacecraft missions to solar system bodies like asteroids, comets, the moon and planets. I was hired by the FBI in 1995 to be an examiner of questioned photographic evidence, basically an expert witness conducting forensic exams of image and video evidence. Since that time, I testified over 60 times in international, federal and state courts as an expert witness. The types of exams I perform cover four areas. Image enhancements like using a fast 48 transform to remove the skin texture obscuring a latent print. So a latent print examiner could see the ridge detail and perform a comparison. I'd also conduct authentication exams. Is it a real image or a fake image? Photogrametric exams. Measuring things from photographs like how tall is the bank robber is another type of exam that we do. And finally, comparison exams. Is the bank robber wearing the same shirt that was found in a suspect's home? Or does the face on this passport really match the face of the person found to be carrying it? Now, a key aspect of this discipline is maintaining an awareness of research developments in image and video analytics to include computer vision. Over the last 14 years, as a senior scientist at the FBI, my primary focus has been identifying new image and video capabilities and integrating them into FBI operations. Those capabilities include things like facial recognition.

Jaclynn McKay [00:03:50] Richard, you have such an exciting background. I'm sure our audience eyes were lighting up just like mine, just listening to you talk about your experience at NASA. I wish we had more time to discuss that, but on today's episode, we're going to stick to the facial recognition side of things. The name is fairly self-explanatory. Can you describe a little bit more in depth about what facial recognition technology is and how it's used to aid investigations?

Dr. Richard Vorder Bruegge [00:04:17] Most people think that facial recognition technology is what you see on TV and in the movies. A computer scans a face, finds landmarks like the corners of the eyes and the tip of the nose, and then takes measurements and compares those measurements to some massive database to locate a 100% match. Well, that's nonsense on a number of levels. First, a measurement based approach, which is called photo anthropometric, or sometimes facial mapping, was proven to be incredibly bad as far back as the 1990s. The best measurement based technique in 1993 was wrong 79% of the time. When asked of two images depicting the same person were given to it. Since then. Pattern based matching is how facial recognition works. Now, pattern based matching refers to the process of comparing the appearance of the face in the given images that are presented to the algorithm. Characteristics like the overall shape of the face. Is it round or oval, for instance, and the variation in bright and dark patches on the face that you might see in the shadows, under the eyebrows and the highlights on the cheeks, as well as the fine texture of the skin, including things like crow's feet, The rougher the skin, the more variable, the texture and the easier it is to tell one face apart from another. Each facial recognition algorithm takes the totality of those characteristic patterns and extracts a numerical representation of those patterns that we call a template. That template can then be compared to other templates, and the degree of similarity or difference between the templates is measured. Now notice how I did not say anything about whether the algorithm is trying to establish the age, the sex or the race of an individual with that template. The process of trying to classify a face based on one of those aspects falls into the category of facial analytics, and that differs completely from facial recognition. Facial recognition is all about comparing one template to the other template and using that result to aid the user in some task. Now, the simplest task you might think of in this scenario is the access control problem. Many listeners may use your face to unlock your mobile phone, and when you first activated the app on your phone, that would allow facial recognition. The app asked you to take a few enrollment photos and maybe even a video. You may have looked straight at the camera and turn your head to the left and right. The app was basically asking you to enroll your face so that it could create a template to compare against. The next time you want to unlock your phone. When you do that, the app is taking a new photo of you live, creating a new template, and then calculating how different that new template is from your old enrollment template. If the difference is small enough, it considers this to be a match and unlocks the phone. If there's too much of a difference, it won't unlock. Many listeners may have had the experience waking up from a night's sleep and find they can't get their phone to unlock. That may be because you're still groggy and your eyes may not be as open as wide as usual. Maybe you have some bags under your eyes that could create enough of a difference in your template that the app doesn't recognize you. This access control scenario is where the field of facial recognition has always had its strongest market. It's now used for identity verification for people crossing borders or going through security at the airport.

Jaclynn McKay [00:07:38] Thank you for explaining all the science behind this technology and for providing our listeners with the real world example of how this is applied in everyday life. You mentioned that using facial templates can be used to unlock phones as well as in other industries such as border crossing. Is this technology used differently at all in the forensics realm as opposed to other industries?

Dr. Richard Vorder Bruegge [00:08:04] The access control scenarios I described are common across many industries in business. Some listeners may recall some controversy last year when the Internal Revenue Service was using facial recognition to verify users identities. In that case, you had a user submitting a template of a driver's license image, for example, and then using a real time camera image of that face for comparison. The driver's license, of course, is a validated identity document, and IRS was using that as a check to allow people to gain access to their accounts. Now, there are a number of business scenarios like that in which facial recognition could be used to assist in verifying identities, which helps them with the provision of remote services. Now, that differs from how we use it in forensics. In forensics. The most common use of facial recognition technology, or a more common use, is a database search. I'm talking about searching a mugshot database images lawfully acquired as part of the booking process and the involvement of a trained human user is critical here. In this case, the user submits a probe image, perhaps a person they're looking to identify, and then a template is created for that probe image. Now, that probe image might be a suspect in a crime like a bank robbery. So you create a template for that probe image. And that template is then compared to the templates of all the images included in the mug shot database. Now, this is sometimes called the gallery, and those results are returned to the user in rank order from most similar to least similar. You can think of this as taking a mug shot book and rearranging the sequence of photos from those. The algorithm rates most similar to those that are the least similar. Now, no human being actually reviews an entire mug book in this case, but the computer can. Once the ordered list is generated, the user, again, who I say is trained to perform this task, reviews a subset, perhaps the top 20 or the top 50 returned images. This is called the candidate list. Now algorithms these days are so good that if you have a mugshot quality image as your probe, that is to say, the image you're investigating. And if a true match is included in the database or gallery, that match will be returned in the top ten and an astonishing percent of the time. 20 years ago, when the latent fingerprint searching was being done, it wasn't unusual for a latent fingerprint search to return a candidate list, and you would only find the match in the top 10 50% of the time. Now, of course, latent fingerprint matching is really good, but when it comes to facial recognition and mug shots, we've achieved that same level of accuracy. Where, as I say, many algorithms can now exceed 99% of the time for that mug shot to mugshot situation. So if a user decides that one of the candidates is a possible match, then they would have generated an investigative lead. It's not an identification, mind you, but an investigative lead. Additional techniques would then be needed to establish the identity of the subject using other techniques. And by that I mean things like finding an eyewitness who knows the subject or finding a credit card record that shows the subject was in the vicinity at the time of the event. You need that added context to support further action like an arrest. Now, we don't always get mug shot quality images to run as probes and degraded images will not succeed at the same rate. They can still do a good job, but that remains an area for improvement. And that happens to be an area of active research right now. And the results of that research can help inform us on how to implement better procedures to avoid false leads. And moving on from the database search. There are other uses of facial recognition that can be incredibly valuable in forensic and investigative settings. For example, facial recognition, along with other video analytics, can be very useful in organizing and triaging large volumes of image and video data acquired during an investigation. In such an application, thousands to maybe hundreds of thousands of images and videos can be scanned to detect the presence of faces, and the user can then be presented with all of those faces one after another. Or you might try to simplify that task by clustering those faces. And by that I mean we try a first pass at grouping together all of the faces that appear to represent a single person. You can also create a gallery of all those faces so that you can search them. So I'm not talking about a mug shot database here. I'm basically talking about a gallery of faces that you have encountered in that set of investigative images. And this is an important difference because in this case, in this investigative gallery, we don't know any information about the person like their biographic data, their name or their date of birth. Instead, we can use those facial images as a way of finding that person elsewhere in the data. And I've got a great example to explain the potential value for you, Jaclynn. Consider the 2013 Boston Marathon bombing. We had about 83,000 still images and 10,000 videos that were either submitted by the public or recovered from the scene or from suspects in the case. Now, this took place around the same time that I was working on the Israel Keyes case. And we have only very primitive triage capabilities at the time. We did not, in fact, have the ability to organize and search that data using the faces in an efficient manner. As a result, hundreds of analysts and investigators took about a year to wade through all that data and tracked the movement of the two Tsarnaev brothers. Today, I'm confident that we would have been able to find 95% of that data in a couple of days and all of it in under a month. Now, one tool that we did have in my lab back then was a clustering tool, and that's what I used in the Israel Keyes case.

Jaclynn McKay [00:14:13] Richard, would you mind telling us more about using the clustering tool in the Israel Keyes case?

Dr. Richard Vorder Bruegge [00:14:18] Sure thing. Israel Keyes was a serial killer who managed to avoid detection for some time because he was meticulous in his preparations and avoided following a pattern, including who his victims were. There was no victim profile. When I say he was meticulous, that's something of an understatement. He would travel around the United States burying or stashing away what he called kill kits. These would include things like zip ties, firearms and silencers. He would leave those in place until he needed to use them for a crime. Now, the beginning of the end for Keyes was when he kidnaped an 18 year old barista named Samantha Koenig from the Anchorage, Alaska coffee Shop in which she worked on February 1st, 2012. By 5 a.m. the next morning, she was dead. Keyes went off on a two week Caribbean cruise with his girlfriend and daughter. After he returned, he created a fake proof of life photo by applying makeup to the corpse, which he had preserved in cold storage in a shed on his property and that day's newspaper. He demanded ransom by having the family put money into the victim's bank account, planning to use her ATM debit card to extract the cash. He then took off to the southwest United States, flying into Las Vegas and getting a rental car. Once he started using the debit card, law enforcement was able to track those uses with the first ping hitting on a bank in Arizona. Closed circuit television images from the ATM showed a masked man with a white Ford focus in the background. Subsequent pings in New Mexico and Texas led investigators to determine that he was heading east on Interstate ten, and a bulletin was set for the white Ford. Texas Highway Patrol troopers caught sight of the car, pulled him over when he exceeded the speed limit and spotted dye stained money in the vehicle. A search revealed Samantha Canning's phone and ATM card. He was arrested and extradited to Alaska. The dye stained money, by the way, came from a bank robbery that he had committed on his way east on I-10. Now, he quickly confessed to killing Samantha Koenig while awaiting trial. He spoke to FBI investigators several times over the next nine months. He cooperated somewhat, declaring that he had killed many people, but he would not identify those other victims except for one couple, Lorraine and William Courier from Vermont. He described kidnaping the couriers from their home, taking them to an abandoned house that was scheduled for demolition. He then tortured and killed them there and deposited their bodies in the basement. The abandoned house was subsequently demolished and all of it was transported to a landfill. Now, despite searching the landfill, the bodies were never found. A fingerprint belonging to Keyes, however, was identified at the courier's home, and the kill kit he had used was located where he told investigators he had stashed it in perishable New York. Now, one other named victim has so far been attributed to Keyes. During the interviews, investigators would show Keyes photographs of missing persons. Most of the time he would say no or nope. But one time he stopped and he said, I don't want to talk about her yet. That victim was Debra Feldman, a suspected sex worker with a drug problem. Beyond her, though, there were no good leads on who else he might have killed. So this is where facial recognition comes into play. A review of the contents of Keyes computers resulted in the discovery of hundreds of photos of Samantha Koenig, the Anchorage barista whose murder started this investigation. A photograph of the couriers was also located on the computer. That's when the Anchorage FBI field office contacted our digital evidence lab and asked if we might be able to use facial recognition to search for other potential victims whose pictures were on his computer. That's how I got involved. Through discussions with the case agents, we decided that a cross matching of the pictures on Keyes computer with pictures from the National Missing and Unidentified Person system NamUs might generate some leads. Now, I earlier mentioned clustering as one of the uses of facial recognition technology. In clustering, a set of facial images is sorted into groups where each group appears to depict the same person and hopefully allows you to reduce the number of subjects to investigate. I used a clustering application that had been built in house for us. The application scanned over 900,000 graphic files from Keyes computer hard drive and located 13,299 facial images that were suitable for facial recognition and clustering. Those 13,000 plus faces ultimately were reduced to approximately 3600 clusters or subjects who appeared to be present on Keyes computer. I then conducted a manual review of those 3600 subjects to locate subjects to run against the missing persons database. You need to realize that Keyes computer contained a lot of photos of himself, his families and his friends, and those photos would be of little value in trying to find missing persons, of course. Now, I was looking for the random subjects who didn't appear to have any direct link to Keyes personal life. The most obvious of these, of course, was Samantha Koenig. The barrister's murder led to his arrest. Her image was present on his computer and hundreds of digital image files, but not necessarily hundreds of unique images, mind you, but hundreds of copy of 5 to 10 different images of Ms. Koenig that appeared to be taken from the Internet. Now, it's important to note at this time that the graphics files recovered from his computer included thousands of images that would have been cached from his Internet browsing history. This could include random pictures from websites he visited. He wasn't responsible for taking 900,000 digital images with a camera so beyond Koenig. After reviewing the 3600 clusters, I identified 520 subjects to search against the NamUs database. Now, at the time, the NamUs database available to me included 21,526 image files, of which 19,186 facial images were suitable for facial recognition searches. So effectively, I ended up running 520 searches, one probe for each cluster against the 19,000 plus images in the NamUs Gallery. Each search generated a candidate list that I then reviewed to see if there was a potential lead. I ultimately found 62 images on Keyes computer whose facial images appeared to correspond to 44 different individuals from the NamUs database, including the courier's. The couple Keyes admitted killing in Vermont. All of these images that they say appeared to be ones that you could find on the NamUs site. And these actually included four subjects that NamUs subsequently removed from their site because the missing persons had either been recovered or their cases otherwise resolved. In other words, Keyes had nothing to do with their being missing. Now, to my knowledge, beyond the couriers and Samantha Koenig, none of the individuals whose pictures were found on Keyes computer were ever conclusively linked to him. So from an investigative standpoint, this effort did not bear a lot of fruit. However, I believe it demonstrated the potential value of this technology for this type of analysis. Let me say one last thing about this particular analysis. The work I just described was performed with an algorithm that is now more than ten years old, making it far worse than what we have available today. In looking back at this case. Earlier this year, I had a chance to search Keyes, facial images using a state of the art algorithm. With such a massive improvement in facial recognition algorithms, you might think that I would have found a lot more images, but you would be wrong. I found a grand total of are you ready for it? One extra image that I had not previously found. It was an additional image representing one of the 44 individuals I had originally identified ten years ago. So there was no additional investigative value coming out of this. That old algorithm, when tested ten years ago, had a rank one accuracy of about 62%. Now, what I mean by a rank one accuracy of 62% is that if you search a gallery of thousands or millions of images, what is the rate at which the true match image comes? Up as the highest scoring match the rank one match at the time. Ten years ago, the leading algorithm would create a rank one return in a range of 92 to 95%. Yet I was using an algorithm that only had a rank one return rate of 62%. Now, today's state of the art algorithms are all well over 99% and returning a rank one match so long as the subject is in the database. Yet from a practical standpoint, the newer algorithm, one that exceeds 99%, would have increased my operational return from 62 images to 63 images. It's not really a massive improvement that one would expect.

Jaclynn McKay [00:23:24] Thank you for describing all your work done in that case. All the numbers of images just sounded overwhelming. But that's really interesting to learn that there was only one image that was missed based on the changes in technology to more current days. Are there any other cases that you'd like to discuss where facial recognition was used?

Dr. Richard Vorder Bruegge [00:23:46] Yes. In fact, I'd say that the Iwo Jima flag case is a really good example to talk about how facial recognition can be used.

Jaclynn McKay [00:23:55] All right, let's hear it.

Dr. Richard Vorder Bruegge [00:23:56] So I was able to assist the Marine Corps in a recent reexamination of the Iwo Jima flag raisers question. And I used facial recognition as a way of testing it in a quote unquote, real world scenario. Now, the Pulitzer Prize winning Iwo Jima flag raising photograph taken by Joe Rosenthal for the Associated Press on February 23rd, 1945, has become the enduring symbol of the United States Marine Corps. You see six Marines straining to raise the stars and stripes with a pole on an angle in the wind whipping the flag. Two days later, when that photo was published on the front page of newspapers across the U.S., the reaction was electric. Here was a symbol of the long struggle to defeat the Axis, demonstrating the perseverance and dedication of our troops to complete the job. It was a sensation. It was so much of a sensation that President Roosevelt decided that he wanted those Marines to be the face of the next war bond campaign. So the Marines have to name them and get them back to the states because they are going to be celebrities. But did they actually know who they were? No one at the time actually considered themselves to be doing anything out of the ordinary. The photographer, Joe Rosenthal, didn't even bother getting the names of the Marines in the photo. This was not uncommon. In this case the Marine command structure had to try to reconstruct who was on the mountaintop that day and who was on the flag based on people's memories. Now, this was over a month after the flag raising. In this environment, through the best efforts available. They did ultimately identify six individuals. Unfortunately, three of the six they identified had already been killed in action. Now, let me pause here and provide another piece of information for why no one thought anything about this second flag raising was important. Is that keyword second flag raising? This wasn't the first flag that had been raised on top of the mountain. The first flag had been raised a couple of hours earlier, and that was of greater significance to the Marines that indicated they had taken that mountain. The second flag was just being put there for better visibility. Also, this is really a good time to stop and remember the full brutality of this battle. This was the highest single action loss in Marine Corps history. It was vicious. It was in this environment that the first American flag was raised on the summit of Mount Suribachi at approximately 10:20 a.m.. That was a huge morale boost because the volcano was the highest point on the south end of the island, and it provided a major tactical advantage to the U.S. by providing visibility over much of the rest of the island. But it was a relatively small flag, only about two feet by three feet. A short time later, a commanding officer on the beach ordered a resupply patrol with the task of replacing the flag with a larger one. Private First Class Rene Gagnon was the runner tasked with taking the flag and batteries up the mountain. Gagnon and several other Marines made their way to the top of the mountain. And at around 12 noon, the first flag was lowered and the second flag raised. Now, remember, no one really considered what they were doing to be any great significance with that second flag. It was a replacement. As a result, even though there were now several individuals on hand with cameras, none of the people taking pictures at that time bothered to get the names of the people who raised the second flag. And as I say, the Marines on Iwa Jima ended up identifying six individuals as the flag raisers from right to left. The Marines identified in 1945 were Sergeant Henry Hansen, who's crouching down at the base of the pole in what is referred to as position number one. Private First Class Gagnon in position two mostly hidden behind the individual in position one. Navy medic John H. Bradley in position three Sergeant Michael shrank in position for barely visible at all behind the others. Private First Class Franklin are saucily in position five and Private First Class Ira H. Hayes at the far left, with his hand reaching for the pole, but not quite on it now, sadly. Sergeants Hanson and Strength, along with PFC Owsley, had been killed in action in the ensuing weeks after the flag raising Navy medic pharmacist mate Second Class John Bradley had also been severely wounded during the battle. But he would survive and subsequently joined PFC Gagnon and Hayes in the war bond effort. Those three not only participated in the war bond effort, but they would also be used as the models for the statue that now stands in Arlington, Virginia. And when the war ended, IRA Hayes tried to get the record corrected by explaining that Sergeant Hansen was not in position one, but that, in fact, Corporal Harlan H. Block was there. This led the Marine Corps to convene a board of inquiry in 1946, which was headed by Major General Pedro del Vallee. They reviewed the matter and in January 1947 issued their findings that Sergeant Hansen had been misidentified and that Corporal Block was indeed the Marine in position number one. And that's how the record stayed for almost 70 years. When the Marines convened another board of inquiry in 2016. This one, led by Lieutenant General Jan C Hooley, retired critical events that led to the 2016 Hooli board reach back to the year 2000, when James Bradley, son of pharmacist, made Second Class John Bradley, publishes a book along with author Ron Powers called Flags of Our Fathers. It became a bestseller and the basis for a Clint Eastwood movie in 2006. Over the next few years after the movie came out, the publicity led to heightened interest in the events at Iwo Jima. And in late 2014, the Omaha World-Herald publishes a story about two history buffs Steven Foley and Eric Krell, who have raised doubts about whether John Bradley was actually one of the flag raisers. These doubts are brought to the Marines who decide there's enough there to convene a new board to investigate. 2016, Hooli board conducted extensive research that included both administrative records of platoon assignments and the like, as well as photographic analysis of the extensive film holdings. The photographic analysis relied a lot on the uniforms, equipment and ordnance or weapons carried by the troops. It's also important to understand that there were multiple photographers on the mountain that day who took numerous photographs and even a 16 millimeter film of the second flag raising. In some cases, individuals depicted in these photographs have been identified. Another critical factor to recognize in this case from the standpoint of forensic analysis is the fact that there were no more than 100 American troops who ever went up the mountain on that day. Given the small set of suspects, differences in uniforms, gear and weapons allows one to more easily differentiate one individual from another and narrow the candidate list down. The 2016 Hooli board ended up determining that Bradley, whose son had written Flags of Our Fathers, was not, in fact in the photograph at all. They determined also that PFC Owsley, who was previously identified in Position three, should be the one who replaced Bradley in position three. Position five, they determined, was held by Private First Class Harold H. Schultz, who had not previously been identified. Two key factors in identifying Shultz were observations from the photographs that differentiated him from all others seen on the mountain that day. He was the only marine documented on the mountain that day, observed to have a broken helmet liner strap. This hanging liner strap is clearly visible in the best photographs we have of Schultz on Mt. Suribachi, including the gung ho photos. The gung ho photos are a critical factor in the flag raising inquiry. This series of photos documents a group of about 18 Marines celebrating the taking of the mountains while they pose in front of the second flag. Most, but not all of the individuals in these images have been definitively identified over the years. And so these gung ho photos provide a set of reference images not just for the individual's faces, but the gear they wear and the ordinance they carry. Another Marine is right in front of PFC Schultz in these gung ho photos, also on one knee, cheering their accomplishment, an individual by the name of Corporal Harold Pete Keller. As we shall see, Keller becomes quite important in the next twist of this story. In the summer of 2018. Amateur historian Dustin Spence. Joined by fellow researchers Steven Foley and Brent west Meyer sent the Marine Corps a 102 slide PowerPoint presentation that offered photographic evidence that Corporal Harold Keller was the Marine in position two not the runner PFC Rene Gagnon. As further support for this proposal. They also suggested that PFC Gagnon could be identified in a second photograph taken at almost the same time as the Rosenthal photograph. That second photograph is focused on the lowering of the first flag with the second flag raising seen in the background. And those photographs were taken by Private Robert R Campbell. So shortly thereafter, Campbell would take another photograph of two Marines saluting the second flag while the first flagpole is pulled away in the lower left hand corner of the frame. The historian suggests that this photo, which has an unobstructed view of the face, uniform and weapon of the Marine in position two offer compelling evidence that it is Corporal Keller and not PFC Gagnon. Now returning to the photo that depicts the lowering of the first flag. The historians would also point out that the individual reaching to grab that first flag appears to be René Gagnon. Based on his gear and facial appearance, the Marine Corps found this evidence so compelling that they convened a new board headed this time by Brigadier General William J. Bowers to address these new claims. It's at this point that our laboratory became involved. They reached out to us explaining the nature of the case and requested our assistance in assessing the work of the historians. We occasionally agree to work cases of historical interest like this, in part because it will allow us to test tools, techniques and approaches that might not be currently validated for use in formal lab work or to stress tests the tools with unusual evidence. Now, let me be clear. Our role was never in any way making new discoveries about these events, but merely assessing the degree to which the claims of the historians could be supported. In particular, the primary questions being asked were, one, could we verify Corporal Keller is the Marine in position two? And two is PFC Gagnon, the Marine reaching for the first flag as it is lowered? Now, my primary contact with the history division was Briann Robertson, who provided electronic copies of the visual evidence they had compiled. So two additional photos that were critical to the historians analysis of Corporal Keller were taken by Army Private First Class George Burns between the raising of the first flag and the second flag. These photos not only provide a full view of the helmet, uniform gear and ordnance carried by both Keller and Snyder, but they are definitely authenticated as depicting them both because Burns made a point of asking for and recording their names and hometowns. These images would also come into play with facial recognition technology later, with all the photos and film evidence in hand. My first task was to treat this as we would any other case in which identification from images was requested. I basically have to review all of the submitted images to find the subjects of interest and locate features that would allow us to include or exclude one person from another in each photo. The Gymnast film was crucial in this process because although it is very low resolution, it allowed me to sync individual high resolution photos to individual frames of the film and then track the movements of individuals associated with each position. In other words, I could prove that the subject in position two in the second flag raising photo was also the subject seen in the background of the Campbell photos, which were focused on the first flag lowering and the Marine saluting the second flag. It was possible to demonstrate that Corporal Keller must be the same individual for a number of reasons. The real capper to this analysis, so to speak, though, is the helmet that he wears. The helmet worn by all of the troops have reversible camouflage cloth covers over a steel shell. Now, it's useful to note there was no effort made to create the same camouflage pattern on every helmet. The camouflage pattern is bold enough that it is more easily seen than smaller details in the images. Now, because of these similarities between the position two subject and the known photos of Keller, including the Burns handshake photos and the gung ho photos, it can be concluded that the subject in position two is indistinguishable from Keller in terms of the uniform, helmet, gear and ordnance. But what about the face? The Campbell photos give one a front view of the position to subject, but the face is in shadow and is not high resolution. Nevertheless, one can clearly see the contours of the face, the chin, the nose and the mouth. One of the Campbell photos shows a slightly darker linear feature extending from the left side of the subject's nose toward the side of his mouth. Contemporaneous photos of Corporal Keller shows that even when he is not smiling, he exhibits a well-defined nasal labial crease, which is what would create the same type of features seen in the Campbell photos. A nasal labial crease is basically a smile line that extends from the side of the nose to just at the side of the mouth. Some people, like Keller have smiled so often that those creases are permanent features of their face, even if they have a neutral expression. So these similarities in gross characteristics of the face lends support for the proposition that Keller is the subject in position two. Now, the forensic science community, of course, has been going through a reexamination of concepts of individualization and identification for many years now, particularly for pattern evidence disciplines like facial comparison or footwear and tire impression evidence. Some of these disciplines are exploring conclusion or opinion scales. This was one area of exploration that I leveraged in this case, basically utilizing a seven step conclusion scale that would extend from minus three to plus three, the plus three being defined as extremely strong support for the proposition that the two subjects are the same, plus two is strong support and plus one is just support or some support for the proposition that the subjects are the same. The negative side mirrors these levels. Now, using this scale, the facial comparison of Keller alone would be a plus one. While the helmet uniform gear and ordnance comparison is a plus three. Taken together, it's my opinion that there is the highest level of support possible for Keller being the subject depicted in Position two in the Rosenthal photo. Looking at Gagnon, on the other hand, we have no authenticated contemporary photos of him on the mountain that day to compare his helmet, uniform and gear. We can, however, compare the face and when one compares the facial features of PFC René Gagnon. There are differences noted in the shape of the nose and the chin, as well as the lack of a notable nasal labial crease, except when he smiles. So when René Gagnon has a neutral expression on his face, unlike Herold Keller, he does not have a prominent nasal labial crease and the individual in position to is not smiling. Based on these differences, I would assign the facial comparison of Gagnon and the subject in position two to a minus one score, i.e. some support for the proposition that different people are depicted. The limited resolution of the Campbell photo just doesn't allow us to be more definitive and go to a minus two or a minus three. One reason we accept cases like this is that they offer a chance to try out new techniques and approaches. Although we use facial recognition technology to search databases for investigative leads, we do not currently use facial recognition technology in our forensic laboratory because it has never been validated for that purpose. Interestingly, the results of a closed box test, sometimes called a black box test, a forensic facial examiners published in 2018, demonstrated that the concept that fusing a state of the art facial recognition algorithm with a trained forensic facial examiners could lead to results comparable to having two examiners perform an analysis and validation. No report leaves our laboratory without having a second qualified examiner verify the results. 2018 research I just mentioned was written by Jonathan Phillips of NIST, and academic colleagues, and was published in the Proceedings of the National Academy of Sciences. It showed that a fusion should be possible, so I wanted to use it in this case to test it. To make a long story short. When I used a state of the art facial recognition algorithm circa 2018 to compare the Campbell photographs to the Burns handshake photographs of Keller, the match scores exceeded the threshold corresponding to a false match rate of 1 in 1000. That was the standard false match rate used for setting thresholds. So in other words, Keller's phone would have unlocked from this facial comparison. On the other hand, when using the same algorithm to compare PFC Rene Gagnon to the subject in position to in these Campbell photos, none of the images known to depict Gagnon come close to the match threshold. The scores are all very low, lower, in fact, than the scores for all other images of Keller's that do not exceed the matched threshold. This highlights the current challenge with incorporating facial recognition algorithms in the forensic laboratory in a 1 to 1 scenario. Exactly the reason I wanted to try it out in this test case. Different images of the same face do not necessarily generate the same match scores. And it's not quite clear what differences in these images lead to different results. When I recently ran these same images through a current modern state of the art 2023 algorithm, none of the Keller images exceeded that algorithm's nominal threshold. But they did return values in the middle of the range. Meanwhile, similar to what happened with the 2018 algorithm, all of the René Gagnon images returned scores that were well below the matching threshold and mostly returned the lowest possible match score. So these results highlighted for me the importance of maintaining a human in the loop for the decision process, which is in fact how we handle facial recognition results in the FBI to this day now, I'd like to close the Iwo Jima story by mentioning one more thing about the seven point conclusion scale and providing a resource available for interested listeners. The Bowers board was very enthusiastic about my use of the seven point scale in conveying my results. They were so enthusiastic, in fact, that they adopted the scale for their use. I ultimately ended up performing similar analyzes for all of the flag raisers, all six positions, and providing results on this scale to assist the Bowers board. They utilize the same scale when delivering their report. Remember, they were looking at far more than just the photographic evidence. Listeners who are interested in learning more about these efforts can download a marine Corps History Division report that's titled Investigating Iwo. It was edited by Dr. Briann Robertson, the individual I mentioned being my primary contact. And it's easy to find with an online search. Just use the term investigating Iwo. It's available for free download as a U.S. government product.

Jaclynn McKay [00:44:37] Thank you so much for those in depth case examples of the use of facial recognition technology. You previously mentioned that there is active research occurring on how to use degraded images or non mugshot quality images in facial recognition technology along those same lines. And in thinking about the future. Where would you like to see this technology go and what aspects need continued growth?

Dr. Richard Vorder Bruegge [00:45:03] Sure thing. As I've tried to convey throughout this podcast, facial recognition technology is an incredibly useful tool when it is used in an appropriate manner. We have been able to help solve a lot of cases and locate a lot of image and video evidence depicting specific individuals using this technology. Image and video analytics like facial recognition allow us to look at all of the evidence evidence that can lead to exonerations, just as it can lead to individuals who should be placed under suspicion. We must retain the ability to triage evidence using these tools. With that in mind, individuals are right to raise concerns about the potential abuse of such powerful technology, as well as the accuracy of it. But they need to understand that agencies such as the FBI put oversight and guardrails in place to limit any such abuse. We don't arrest a person just because a facial recognition algorithm had a high match score to them in a gallery search. We require investigators to develop additional evidence to support further law enforcement action. I want to highlight the need to continue to conduct research in this space. NIST testing, National Institute of Standards and Technology shows that for passport quality photos, there are many algorithms that have performance well above 99% accuracy. The current challenge we face is in the area of lower quality images. We are supporting research in this space in an effort to see how the performance of algorithms changes with differences in the size of the image and the blurriness of the image. And that, I think, is where we we need to be looking at the technology today.

Jaclynn McKay [00:46:42] Thank you, Richard. Those are great points to end on and I appreciate the call to action for further research. It has truly been a pleasure chatting with you today and thank you so much for your time.

Dr. Richard Vorder Bruegge [00:46:53] It's been my pleasure, Jaclynn. Thank you very much.

Jaclynn McKay [00:46:56] If you enjoyed today's episode, be sure to like and follow just science on your platform of choice. For more information on today's topic and resources in the forensics field, visit forensics. COE dot org. I'm Jaclynn McKay, and this has been another episode of Just Science.

Introduction [00:47:17] Next week, Just Science sits down with Ross Krewenka and Mike Ransom to discuss the rise of contactless fingerprint technology and its impact on the future of ID. Opinions are points of views expressed in this podcast represent a consensus of the authors and do not necessarily represent the official position or policies of its funding.