Law Enforcement Applications of Forensic Face Recognition

MICHAEL PETROV, PHD
Director, Advanced Solutions
Contents

3 WHY FACE RECOGNITION?
3 RECENT SUCCESSES
4 FORENSIC SEARCH PROCESS
5 WHAT IMAGES CAN BE USED FOR FACE RECOGNITION?
7 BRIEF HISTORY OF FACE RECOGNITION ALGORITHMS
10 OBJECTIVE ASSESSMENT OF ALGORITHM ACCURACY
12 CRIME SCENE IMAGE ENHANCEMENTS
14 THE LEGAL ASPECT
15 FACE RECOGNITION SYSTEM ARCHITECTURE
16 CONCLUSION
Why Face Recognition?

Some law enforcement agencies estimate that up to a quarter of complaint cases contain face images of the suspect or an accomplice. This number is significantly higher than for latent fingerprints or DNA samples. Crime scene images arrive from a variety of sources including social networks, CCTV recordings, documents, cell phones and other media. While the images are available to the detective, they are still rarely used for investigative purposes. Law enforcement agencies are just beginning to adopt face recognition technology to enable their officers to search for these images in mugshot databases.

Investigative face recognition technology has already been proven to be highly efficient at a number of agencies. Face recognition searches take just seconds and produce results that are easy to visually adjudicate. The searchable databases available to law enforcement agencies achieve unparalleled population penetration coverage, because in addition to arrest and arraignment databases, they may in certain jurisdictions include DMV and other databases.

This white paper shares our recent experience of deploying face forensic solutions at several large law enforcement jurisdictions. Our solution architecture approach has evolved to a system that is tightly integrated with the agency’s business processes, including case management and booking, while the sophisticated investigator software is capable of extracting the most from the crime scene information using face image preprocessing and efficient search methodology.

Recent Successes

Recent successes in face recognition applications in law enforcement may be attributed to the following factors:

1. Dramatic increase in the accuracy of commercial algorithms provided by the top tier vendors. Incremental improvements lead to an order of magnitude accuracy improvement every four years as demonstrated in NIST tests.

2. Standardization of the mugshot capture process across most agencies leads to better quality of databases and encourages inter-agency database exchange.

3. System architecture and examination process maturation. While the early systems were essentially “face lookup” applications, the users of the new generation systems are trained facial examiners who understand the business process of the organization and use sophisticated software tools.
4. Proliferation of social networks where users upload their photographs.
5. Improvements of photographic quality of CCTV and consumer digital equipment.

All these factors change the way face recognition is used in law enforcement. Document-type photograph searches constituted a typical success story of the first-generation deployments. Compare this with typical success stories of the new generation of systems. A woman had an unfortunate encounter with a stranger whom she met on a dating website. The perpetrator’s name and other personal information on his social network page were intentionally deceptive, but the photograph was genuine because his intent was to eventually meet the victim in person. Biometric search of the dating website profile photograph produced an accurate match that led to arrest.

In another example of recent technology use, the CCTV imagery retrieved from a crime scene contained the face of the perpetrator. However, the quality of the images was too low to allow accurate unrestricted search against the entire image database. The face examiner used her crime analytics expertise to restrict the biometric search primarily, though not exclusively, to the criminals of a particular race, gender, age group, prior charges, and even the location of prior bookings. The search produced two leads which were further investigated by detectives in the field, with one of the leads leading to arrest.

Many federal, state and local investigative agencies currently count dozens or even hundreds of success cases.

**Forensic Search Process**

Once the forensic image is uploaded, the automated face recognition search takes several seconds and produces a list of candidates as shown in Figure 1. The images are ranked in order of their algorithmic similarity. Depending on the image quality, a substantial number of candidates may need to be reviewed by the examiner during the adjudication process. However, unlike that of latent fingerprints, the manual adjudication of face images is very fast due to the natural ability of the human brain to recognize faces. Even a list of several hundred candidates can be adjudicated in less than a minute.
In addition, the search software provides the capability of processing multiple images of the same suspect as a single transaction and reviewing the results in a fused sorted view. Prior to searching, low quality images may be enhanced as described in this white paper.

The use of demographic filtering to make searches more accurate is important in latent fingerprint systems, but it is crucial for face forensics. Unlike fingerprints, photographs of faces instantly reveal substantial filtering information, even though it should be used with caution due to possible variations of the subject’s appearance and accuracy of data. Gender, race and approximate age can usually be determined with high confidence. Height range of subjects on CCTV images can sometimes be established by visiting the site and comparing the recorded appearance of the perpetrator with the appearance of a person of known height and thus may be accurately determined too. Complexion, hair style and length, eye color, facial jewelry, and other characteristics are low confidence filters and thus should be used with caution. We have analyzed the accuracy of the demographic data based on the statistics of subsequent arrests at one of our customer’s databases. For example, the gender descriptor error rate appears to be much less than 1%; the race is consistent to within 3%, age in the ten-year range is 3% accurate, while hair color is only accurate to within 30%.

Additional demographic filtering can be applied by experienced examiners in the context of the crime description. For example, certain crimes may be localized to prior arrests in the same and nearby precincts.

What Images Can Be Used for Face Recognition?

Crime scene images are collected from a growing variety of sources, such as CCTV systems, cell phones, social networks and documents. A noticeable trend is that the amount, quality and resolution of digital images collected at a crime scene increase every year. As an example of resolution improvements, a recent Nokia cell phone boasts a resolution of 41 megapixels. CCTV systems are also developing in the direction of higher resolution and frame rate. High quality photographs obtained from a fraudulent document are almost certain to produce hits. Photographs from other sources may or may not produce hits depending on their quality and context.

Social networks are currently an essential crime solving mechanism in the detective’s tool chest, but their importance for face recognition applications is unprecedented. Not only do these networks contain numerous photographs, but even though perpetrators frequently provide deceptive personal information they still upload their photos into their profiles or
shared images. On average, the quality and thus the likelihood of getting a hit are very high.

CCTV archives are good sources for face images. The majority of modern CCTV systems use DVR recorders in the back end. The videos are recorded in MPEG or similar digital formats. A new trend used by some chain stores and franchises is not to keep local DVR recordings on premises, but instead to upload the digital video streams to a centralized secure data center via the Internet. The advances in digital CCTV technology make it easy for a police detective to retrieve a video snippet recorded at the crime scene. Since different DVR recording brands use different coded software for storing the files, the face examiner needs to collect a number of plugins to be able to play back video files from across the jurisdiction.

Digital CCTV systems allow the end user to configure system settings. This may create inconsistency of video quality across locations even when they use the same cameras. A DVR can be set to record a low frame rate — for example, 1 frame every 3 seconds — to conserve disk space, and as a result the face of the suspect may only appear on a couple of video frames as shown in Figure 2. While some of these images may still be used for forensic searches, the best results are achieved when the face appears on multiple frames with several of the frames capturing frontal face images of good resolution and no occlusions.
Several examples of suitable images are provided in Figure 3 below, while Figure 4 gives examples of images which are not likely to produce hits. Even if a list of likely candidates is produced for these images, the examiner may not be able to perform visual confirmation against mugshots with reasonable confidence. (All the displayed example images are from the public media domain and not connected to actual face investigation cases.)

**Brief History of Face Recognition Algorithms**

The birth of face recognition as a computer application discipline goes back to the mid-90s when several algorithms were proposed. Remarkably, all
modern commercial algorithms, except for the skin texture analysis, can be traced back to university research during that period.

Of the original algorithm families, only two have survived the tests of time. Others, such as initial textbook implementations of Eigenface algorithm, though promising in the beginning, have been unable to scale in terms of accuracy and tolerance to variations in lighting and pose, and have thus been abandoned by commercial providers.

The first of the mentioned above algorithms was developed by researchers from Joseph J. Atick's laboratory at Rockefeller University in New York. The seminal paper by Penio S. Penev and Dr. Atick is called Local Feature Analysis: A general statistical theory for object representation (Computation in Neural Systems 7 (1996), 477–500). Figure 5 illustrates the idea behind the method. The face image is first normalized in terms of lighting and landmark locations, and then split into a several dozen regions. Then the regions are scaled, further transformed and compared pixel-by-pixel. The more similar the areas are, the higher the matching score would be and thus the higher the probability that the two images belong to the same individual. In early US government vendor testing, called FERET (Facial Recognition Technology), this approach was shown to outperform all other algorithms of that period in terms of accuracy. Under the original name Local Feature Analysis (LFA) the algorithm has been actively developed by MorphoTrust USA and its predecessors Visionics, Identix and L-1 Identity Solutions in their New Jersey office, with Dr. Atick still involved in providing the strategic direction of the technology development. More recently, NEC in Japan commercialized a similar algorithm with one of the patents attributed to Dr. Penev, Dr. Atick's graduate student at Rockefeller and co-author of the ground-breaking paper on local feature analysis.
The second algorithm was influenced by the ideas of the researchers of Prof. von der Malsburg group at Rurh-University in Bochum, Germany, first expressed in a paper by L. Wiskott, J. M. Fellous, N. Kuiger and C. von der Malsburg entitled Face recognition by elastic bunch graph matching (IEEE PAMI 19 (1997), 775–779). Prof. von der Malsburg’s colleagues and students had an impressive impact on the industry commercializing the method through a number of businesses including Viisage/ZN (now part of the MorphoTrust algorithms), Cognitec Systems, Neven Vision (now part of Google/ Picasa and founded by Hartmut Neven, a graduate student of Prof. von der Malsburg), C-VIS (now part of Cross Match). Originally called Elastic Bunch Graph Matching, variants of the algorithm currently have a number of marketing names including Hierarchical Graph Matching (HGM) at MorphoTrust. Figure 6 illustrates the basic principle. First, a graph map is fitted on the normalized landmarks of the face, and landmark signatures are then computed using wavelet transformation and compared between the two images.

With all commercial algorithms traceable to just one of two original methods, the only exception is the more recent skin texture recognition algorithm, patented by MorphoTrust, that compares minute skin wrinkles, pores, scars and other artifacts, patented by MorphoTrust. This algorithm is still practically not applicable to face forensics applications since it requires mugshot resolution of crime scene images which is rarely available. However, with the ever-increasing resolution of commercial video cameras this algorithm may have applications in the future. The unique feature of the skin algorithm is that an image of only a small patch of the face skin, amounting to roughly10% of the total face skin area, may be sufficient for identification.
The major vendors have spent years incrementally perfecting the algorithms to overcome challenges presented by low resolution, suboptimal pose or expression, non-uniform lighting, and other limitations. The improvements apply not only to the core matching algorithm, but also to the automatic preprocessing of images prior to template creation. Figure 7 gives an example of image enhancement performed automatically before the image is processed by the main algorithm: the landmarks of the face are located, a 3D mesh is applied, rotated, and projected back to the normal pose view, and then lighting is equalized across the image.

In spite of all the enhancements, conceptually, the algorithms retain many of their original benefits and drawbacks, and thus the algorithms within the two main families are highly correlated; the search results will overlap to a large degree from similar types of algorithms from different vendors. However, the algorithms from two different families correlate much less.

Realizing the complementary nature of LFA and HGM algorithm families, the ABIS® search engine from MorphoTrust applies both in a sequential search process. The search is performed in two or more passes: first HGM and then LFA, with the results fused. Importantly, the fusion is done at the basic algorithm feature level using the scores normalized to the overall anticipated algorithm performance. This algorithmic flexibility is unique in the industry.
Objective Assessment of Algorithm Accuracy

The U.S. government has been involved in biometric vendor testing since the earliest days of the industry to establish objective metrics for procuring the technology. The National Institute of Standards (NIST) is mandated to perform the tests in practical scenarios of face recognition applications at government agencies. These test scenarios include comparisons of mugshot, low quality image, video, and 3D files. The current NIST program sponsors are the Department of Homeland Security (DHS), Department of Justice (DOJ), the Office of the Director of National Intelligence, and the antiterrorist interagency group Technical Support Working Group (TSWG) providing funding and their databases for the tests. The objective scientific approach makes NIST tests the gold standard of accuracy assessment in the U.S. and the entire world. The algorithms from all major vendors have been carefully benchmarked with tabulated accuracy parameters available in public domain at the face.nist.gov/frgc/ website and other publications.

The NIST tests allow tracking the progress of the industry since its inception. Figure 8, from the most recent report, demonstrates the dramatic reduction of error rates and thus increase of overall accuracy over time. Note that the accuracy has improved by a factor of 8 during the 2002-2006 time period and then again by a factor of 9 from 2006 to 2010. During the last decade,
face recognition research methodology has matured, sizable training databases have been collected, and the feedback from practical deployments has been incorporated, which has led to a dramatic accuracy jump.

Not only has face recognition become more accurate, but it has also become faster and more computationally efficient. Currently, the hardware footprint of a face recognition server may be 10 times smaller than that of a fingerprint system with the same number of personal records and search throughput. A single server computer can contain over 10 million records with search time of less than 10 seconds.

MorphoTrust and its predecessor companies Visionics, Identix, Viisage, and L-1 Identity Solutions have participated in the NIST tests from their inception. In every single test, the company’s technology was ranked as number one or number two. In the most recent test, NIST MBE 2010, which technically occurred in 2011, MorphoTrust was number one in accuracy in searching large databases (over 1 million records) and number two in searching smaller databases.

The progress in biometric platform architecture has been equally impressive. The best search recognition engines perform multi-biometric (face, fingerprint, and iris) searches in the same software application, thus reducing system complexity. They run on commercial off-the-shelf (COTS) server hardware and use commercial databases for data storage. The NIST Electronic Biometric Transmission Specification (EBTS) standard is used for input/output transactions to seamlessly integrate with other law enforcement IT systems.

Crime Scene Image Enhancements

Face recognition search matches images against frontal-pose mugshot images stored in a police database. Even when the quality is low, the forensic photographs may still produce hits with the help of image enhancement tools. Contrary to common perception, most traditional photo enhancement tools do not help face recognition. For example, changes in brightness, contrast, color balance, or application of a sharpening mask may improve visual appearance of an image but deteriorate face recognition accuracy. The main reason for that is the extensive pre-processing is already performed automatically within the face recognition algorithm itself, while any manual image transformation leads to the loss of original image information.
MorphoTrust has developed a number of image enhancement tools for improving face recognition performance. These proven tools are part of the MorphoTrust Face Examiner Workstation (FEW) software application that is the front end for forensic searches. Low quality images are never guaranteed to produce a hit. However, the tools improve the statistical odds of getting a hit.

The pose compensation tool, as shown in Figure 9 rotates the face back to the normal pose if the face has been photographed at an angle. It fits an implicit 3D mesh model under the texture of the photograph, then rotates the 3D mesh, and renders the texture back to produce a flat image.

Figure 10 demonstrates how the illumination enhancement tool equalizes asymmetric face lighting and removes hot spots.
Application of de-interlacing tools is shown in Figure 11. Interlacing is one of the main factors deteriorating the quality of stills acquired from analog video signal. There are several commercial tools available for removing interlacing artifacts or de-interlacing. The de-interlacing algorithm developed by MorphoTrust is designed to work with face images thus providing superior accuracy.

Single Image Resolution Enhancement (SIRE) is a unique image enhancement tool designed for face recognition. This algorithm removes blemishes on a low resolution image and increases the effective resolution as shown in Figure 12. The algorithm is trained on a large database of human faces, and it reconstructs missing information by assuming that the object in consideration is a human face.
The Legal Aspect

The results of automatic face recognition searches are normally used as investigative leads. Generation of investigative leads by means of automatic computer algorithms is a conventional practice in law enforcement. The written report of a forensic examiner normally indicates that the biometric hit is not the direct reason for arrest. Court presentations of 1:1 face identification results are still rare. At these presentations, the expert testifies whether the images belong or, more often, don’t belong to the same subject. Even though the use of facial features for subject identification goes back over a 100 years (see Figure 13), many types of face comparisons still do not satisfy the Daubert standard of court admissibility. While the traditional anthropometric face measurements are well understood and tabulated to prove their uniqueness, they cannot be performed on low quality photographs, the area of current interest.

The work to satisfy Daubert admissibility criteria for court testimonies of face recognition experts is a current topic of biometric research. The research is focused on determining an associated error rate of the identification technique.

In late 2011, the FBI started providing a week-long training to professional face investigators that covers a number of relevant topics including face aging, bone structure, and anatomy. The U.S.-based Facial Identification Scientific Working Group (FISWG) is an inter-agency association that has a charter to develop face comparison processes. The eventual goal is to create a facial examiner certification program similar to that of fingerprint examiners. Among participants of FISWG are the FBI, DoD, DOS, and NYPD, as well as several international agencies such as the London Metropolitan Police.
In general, the four methods of facial 1:1 identification, frequently used in combination, are the following:

1. Non-quantifiable Holistic Comparison, where conclusions are reached by visually comparing images as a whole.

2. Superposition, where images are scaled to the same dimensions and overlaid. Though practiced by some jurisdictions, this method does not appear to have scientific backing. In fact, the outcome of a superposition analysis is highly sensitive to the subject’s pose and thus may easily produce inaccurate results.

3. Morphological Analysis, where individual face features are compared and classified. This method is the most applicable to modern face identification.

4. Anthropometric Analysis, an extension of the established discipline of spatial measurements of facial features, and distances and angles between facial landmarks.

In support of Daubert admissibility, researchers in the U.S. and Europe are working to establish the probabilistic occurrences of different morphological features such as moles, wrinkles, scars, shape of the mouth, eyes, and chin that can be observed on crime scene photographs.

**Face Recognition System Architecture**

How to organize forensic operations is an important early system planning decision.

The early face recognition solutions were essentially “face lookup” services available on the agency’s website to allow any detective with proper login to perform a search. JNET, the first system in Pennsylvania, is a good example, with over 500 trained users across the state. The agency provides basic training while leaving the investigative process to detectives and other users in the field. The wide user base promotes system use. However, since the detectives use the system only occasionally, the majority of them do not acquire the breadth of domain expertise required for performing sophisticated searches such as those involving low quality images where different system settings may need to be kept fresh in mind. The problem is not only technical. Recent research by, for example, Michael Bromby and others in the United Kingdom confirms the notion that professional examiners do a much more accurate job performing visual adjudication of potential hits than occasional users. Since the professional examiners are able to adequately familiarize themselves with the database content, algorithm behavior and the crime scene images, they know what to focus on during the searches.
As the applications mature, face searching becomes a forensic discipline where trained and certified examiners perform their work using a variety of complex software instruments. Modern face forensics is organized as a lab which is integrated with the business processes of a police organization. Figure 14 shows an example of the MorphoTrust solution architecture consisting of the following components:

1. ABIS face recognition search engine, the essential building block. The search engine may contain from thousands to tens of millions of biometric templates and demographic binning profiles. The images, including the unsearchable SMT (scars, marks and tattoos) and profile images and full biometric text descriptions, reside in a separate database that is integrated with the search engine core.

2. Face Examiner Workstation client software, a sophisticated client application for performing searches.

3. ABIS Workflow server integrates the system with the arrest and arraignment process and the case management system using transactions based on Electronic Biometric Transmission Specification (EBTS) standards.

4. Biometric Identification (BI) web application that provides basic investigative search and mugshot record management functions, such as generation of lineup, and supports mobile identification.
The system operates as follows.

1. A detective collects crime scene images and submits a complaint report.

2. The case management system transfers the request to the ABIS Workflow server, which audits the transaction and places it in a work queue.

3. The forensic examiner accesses the queue using the Face Examiner Workstation software, performs searches in the ABIS database, generates a search report and then submits the transaction back to the Workflow server.

4. The case management system attaches the report to the case complaint form and communicates it back to the detective.

5. The detective receives the search report and completes investigation in accordance with the police department procedures.

If the forensic image does not produce a hit, it may be enrolled into the Unsolved Crime Photo database of the ABIS server. Every new booking record will subsequently be searched against that database, and if there is a possible hit, an operator adjudication alarm is sent. This Unsolved database may also contain FBI Wanted lists or Terrorist Screening Center watchlists.

Another workflow is used for enrolling new booking records. During the ongoing enrollment process, the ABIS Workflow server evaluates the quality of the mugshots in the booking records the metrics specified in the ANSI/International Committee for Information Technology Standards (INCITS) 385-2004 Face Recognition Format for Data Interchange face image quality standard. If the mugshot is not satisfactory, then it is placed in a queue for quality review by the forensic examiner on the Face Examiner Workstation. This process is designed to prevent enrolling low quality mugshots and non-faces into the ABIS database.

Integration of the search process into a case management system enables adequate control and audit of all search transactions. It also ensures the completion of the investigative loop as the detective investigates and confirms the leads. The audit logs of the case management system can be analyzed to determine system performance.
Conclusion

A recent news article reads: “…detectives have used cutting-edge facial recognition software to capture a man suspected in a shooting at a barbershop. The bullet grazed the 39-year-old man’s head. He required stitches, but will survive. The victim knew the shooter but not his name, pulled up a Facebook photo and gave it to police. Authorities at the Real Time Crime Center fed the photo through the system and a match appeared: The suspect’s prior mug shot.” This white paper explains what happens behind the scenes.

Face recognition is becoming an indispensable tool for modern-day law enforcement agencies. Whether inside the lab or the squad room, face recognition is continuing to prove itself as the next generation tool for identity verification. The efficiency of the deployment depends on the accuracy of core algorithms and on how well the technology is integrated into the business processes of law enforcement agencies.

About the Author

Dr. Michael Petrov designs and implements innovative biometric systems in his capacity as Director of the Advanced Solutions Group of MorphoTrust USA. In his earlier role as R&D manager at L-1 Identity Solutions, Michael led the development of the current real-time face screening Argus product and conducted research in experimental biometrics. Michael holds a Ph.D. from Princeton University.