Chapter 10
Proof of Identity

Written records and verbal testimony are fundamentally fallible: documents can be forged, and people lie. Therefore, these cannot be relied upon alone to conclusively prove one's identity. This poses a difficult problem for institutions mandated with identity management. Biometrics is an appealing solution, as it provides a link between an individual and who they claim to be. This is not the panacea for all problems at hand, but will undoubtedly play an momentous role in the future of identity management and criminal investigations. This has important implications for the use of biometric data in a commercial and legal setting, which is the subject of this chapter.

This chapter is comprised of two sections:

- With the growing prevalence of identity theft, it is apparent that identity management systems, such as those pertaining to driver's license and passports, need reliable procedures for the issuance of new identity documents. The first section presents the practical ways that biometrics can be used to strengthen the issuance process, and develops novel methods for assessing the level of fraud that has infiltrated existing systems (Sect. 10.1).
- The second section considers the use of biometrics in a legal setting. This is an emerging area, and the output of biometric recognition systems is yet to reach a stage where it is fully embraced by the courts. However, there are reasons to believe that it will begin to play a greater role in future legal cases. A comprehensive treatment of the subject is not presented here, but rather a commentary on how the ideas that form the core of this book are vital to the concept of 'proving identity' (Sect. 10.2).

10.1 Identity Document Systems

Identity theft is a type of fraud that is committed by falsely assuming another person's identity, and generally involves the acquisition of ID documents under the victim's name. Once a false identity has been established, it can be used to obtain money or other benefits that the victim is entitled to. For example, if a criminal has sufficient identification under another person's name, he or she may be able to obtain credit cards or loans using the victim's line of credit. The financial and emotional damage can be severe, and it often takes a considerable amount of time and money for the victim to reestablish themselves. It is estimated that identity crime was valued at over US$55 billion in 2006, and the trend is growing worldwide. With so much money at stake, identity theft has become one of the main interests of international organized crime groups. Consequently, criminals are becoming increasingly opportunistic, devious, and sophisticated in their efforts.

The serious nature of identity crimes has exposed the key role of governments and Proof of Identity (POI) issuance agencies in strategies to prevent identity theft and fraud. The first line of defense is stricter procedures to ensure that agencies only issue identity documents to the rightful owner.

Almost all driver's license and passport documents have an associated photograph of the individual, and enormous databases of these photographs are maintained by the issuing body. A facial photograph is a biometric sample, and therefore provides a link between a photograph and the person who was photographed. Assuming that the initial acquisition of a document was legitimate, this can be used to help determine the authenticity of all subsequent applications. Therefore, there is intense interest in the use of biometric matching to contribute to a more robust POI issuance process. Furthermore, there is increasing interest in embedding additional biometrics, such as fingerprints or irises, in ID cards or passports for stronger authentication in the future.

It must be kept in mind that the aforementioned link between a photograph and the person photographed is by nature probabilistic. In other words, the matching process is not deterministic, and there is an inherent degree of uncertainty in any biometric decision (see Chap. 7). Therefore, all businesses rules and processes regarding the adoption of biometric technology must be appropriately grounded in statistical reasoning. With this in mind, Sect. 10.1.1 examines the most appropriate ways to employ the use of biometric technology. The general focus is on the use of facial biometrics by driver's license and passport authorities. However, the concepts are applicable to any organization that maintains large identity databases, regardless of the agency or biometric modality. Section 10.1.2 examines the problem of estimating levels of fraud in existing systems.

10.1.1 Modes of Operation

The primary interest in the use of biometrics in the context of identity management stems from its ability to help reduce fraud. For example, consider a typical driver's license authority. Due to its high penetration rate, driver's licenses have become a de facto "proof of identity" in many countries. Federal and state government, law enforcement, utility companies, and financial institutions rely heavily on the integrity of the driver's license for their proof of identity business processes. Therefore, improving the ability to resist and reduce fraud will have significant flow on benefits to governments and the business community at large.

There are various ways in which biometrics can be used to strengthen the POI issuance process. In general, the approaches can be divided into front of house techniques and back of house techniques. Front of house processes are conducted at the time of submission, while the applicant is on-site. A biometric decision regarding identity is made in real-time, and the result is generally presented to an operator who takes appropriate action. Back of house processes are periodically run from a central processing facility in batch mode. For example, all of the new applications for a given day may be processed overnight, with the results reviewed by dedicated investigators every morning.

10.1.1.1 Front of House

There are two basic ways that biometric matches are conducted: verification (Sect. 7.1) and identification (Sect. 7.2). Verification is a one-to-one match, and seeks to answer the question "Is this person who they say they are?". Identification involves a one-to-many match, and addresses the question "Who is this person?". In theory, either method can be used in a front of house manner when someone is applying for a new ID. However, there are limitations to the identification approach, which will be outlined below.

Verification

In the case that an ID document is being renewed, there will often be an existing photo available from the previous application. Verification can be used to help validate that the person attempting to acquire the new ID is in fact the same person who enrolled previously.

The traditional, non-biometric approach to validation is to display the previous enrollment image to an operator, who performs a visual comparison between it and the person standing in front of them. There are two problems with this approach. The first problem is known as *operator fatigue*. An operator may serve over a hundred clients in a single day, and the a priori probability of a fraudulent application is generally very small. For example, it is likely to be significantly less than 1%. Therefore, almost all of the images presented to the operator will be a genuine match. Over time, the operator will likely become tired of closely examining every image when true cases of fraud are so rare. Furthermore, they may be hesitant to inconvenience people when they are uncertain in their own mind. Eventually, they will become habituated to hit "accept" after a brief, cursory glance, adding little security to the issuance process. The second reason is that even when full attention is being paid, the verification task is very difficult. In general, and operator performance tends to be overestimated. The reality is that for most people the human visual system has a surprisingly high error rate when presented with two unfamiliar faces. Therefore, one should not rely on operators as the sole means to prevent fraudulent applications.

An obvious and natural application of biometric technology is to match a new, high-quality image of the applicant against the previous enrollment. A decision can be made based on the similarity score on whether or not to alert the operator of a potential case of fraud. In comparison with a solely operator-based approach there are two advantages. First of all, the system will operate very quickly, will be consistent, and will not be subject to fatigue. Secondly, although performance is not perfect, at least the algorithmic weaknesses can be evaluated and quantified (see Chaps. 7 - 9).

From a biometric point of view, the process is as follows. In the absence of fraud and data labeling errors, all matches will be genuine as the applicant is a correct match with himself or herself. Therefore, when a particularly low score is achieved it indicates that the applicant may not match the image on file. A match threshold is fixed, and when a verification scores below this threshold, the operator is alerted of a possible case of fraud.

In order to evaluate this scenario, the error rates of most relevance are verification performance measures of Sect. 7.1. The following are the two error rates of primary interest:

- **False non-match rate:** This is the probability that a legitimate transaction is rejected. In other words, a person is trying to renew their ID, and the system falsely suspects a case of fraud. This will occur for low scoring genuine matches. Ideally, the match threshold should be set low enough that this outcome is rare.
- **False match rate:** This is the probability that a fraudster will not be detected. In other words, a person is attempting to acquire an ID under an assumed identity, and the system does not alert the operator of a likely case of fraud. In this case, the match is an impostor match which has achieved an unusually high score. Ideally, the match threshold is set at an operating level that is high enough that impostor matches rarely exceed it.

As can be seen from these error rates, there is a trade-off between false matches and false non-matches. One of the primary goals of an evaluation is to estimate these error rates in order to help find an acceptable operating point. The ROC curve (Sect. 7.1.3.1) expresses these error rates over a range of threshold values, and is therefore an integral part of the evaluation.

Recall that error rates are based directly on genuine and impostor score distributions. The genuine distribution determines false non-match information and the impostor distribution determines false match information. The accuracy of these rates relies on the implicit assumption that the distributions were built using the same type of data that is being evaluated. Therefore, building these distributions must be done with care in order to ensure the relevance of the results.

Building a genuine distribution is relatively easy, as it only involves selecting a range of images from the same person. The participants should be selected from a random cross-section of the population. The data quality of both the enrollment image and the verification image should reflect the data quality of the real enrollment and verification images, ideally captured using the actual equipment. Another factor to take into consideration is template aging. It is important that the procedure for selecting images for the trial is not biased to select images separated by a constant length of time. Rather, a full range of time, from months to years, should separate the images.

Computing the impostor distribution generally requires more attention, and depends on the intended application. For front of house verification, the impostor distribution is used to estimate the probability that a person is not who they claim to be. However, fraudsters will not pick their victim at random. For example, a young Caucasian male would not be wise to walk into a driver's license office and try to obtain a license belonging to an elderly Asian female. This information should be made implicit in the impostor distribution by only matching people from similar demographics. In other words, building an impostor distribution by matching people at random will not accurately reflect the match scores resulting from people actually committing fraud.

Operator Reaction

The question of what action an operator should take when a likely case of fraud has been detected is important, but difficult. In general, a replacement license should not be issued on the spot. However, it should also be kept in mind that biometric matching will inevitably make mistakes, so applicants should be presumed innocent pending further investigation.

Operators and Match Scores

Another operator-related issue is the presentation of results. It is generally not advisable to present operators with raw similarity scores. This is because people without a background in biometric analysis will often invent their own interpretations of the score values. For example, people are naturally inclined to interpret any score in the range 0 to 100 as a probability estimate of a correct match. For example, a score of 99 would likely be interpreted as "there is a 99% chance that these are the same people". However, statements such as these are rarely correct. Therefore, there is a case to be made for not presenting numerical results to an operator at all. Rather, one can use a small number of well defined categories, such as "Unlikely Match", "Uncertain" and "Likely Match".

Identification

In theory, biometric matching can be used to determine if a new applicant already exists in a database under a different name. For example, a front of house system could be designed to return all likely matches for a new applicant from the database. There are two approaches for candidate selection: returning all matches above a predefined match score threshold, or always returning the top *X* ranked matches.

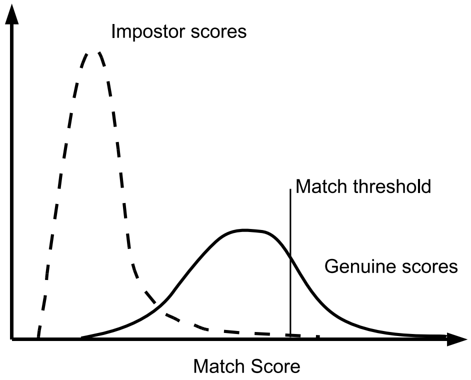

Fig. 10.1 Genuine and impostor score distributions

A threshold based approach is generally used for open-set identification, where it is unknown whether or not the query person exists in the database. This is the case for fraud detection, where actual cases of fraud are rare. Unfortunately, for large databases there will be a large number of impostor matches with high scores. This can be illustrated using the match score distributions of Fig. 10.1. First consider the impostor score distribution. A match threshold needs to be selected such that very few true impostor matches score above it. An example of such a match threshold is indicated by a vertical line. Assume that the false match rate at this point is 0.00001 (or 0.001%). For an identification query, the input is matched against every database enrollment. Consider a database with 1,000,000 images. For a given input, the expected number of false matches is $0.00001 \times 1{,}000{,}000 = 10$. Now consider the identification rate at this point, which is the area under the genuine curve, to the right of the match threshold. For the example of Fig. 10.1, the probability of a correct match looks to be around 30%. Therefore, for any given applicant, an operator would expect to see 10 false matches, and in the rare case of a fraudulent application, there would only be a 30% chance of the correct result being returned. Of course, these number depend on the specifics of the database size and quality, and the matching algorithm being used. However, these numbers are representative of what can be achieve by the best systems of today, and are unlikely to be considered adequate performance.

A rank based approach does not fare much better. Recall that a rank based approach always returns the top *X* ranked matches. The actual cases of fraud will be rare, so operator fatigue will be a major factor. The actual enrollments returned will be the closest matches from a very large database. In general, with a large dataset there will always be a few people who look similar to any given applicant, and this makes the recognition task for the operator very difficult. Furthermore, even if an operator is willing to look at the top 20 matches, the identification rate for large databases may still be fairly low.

In summary, with a large database there will always be people with coincidental similarities, so full-scale front of house identification is generally not feasible. This is another example of the dependence of identification performance on database size, as explored in Sect. 7.2.1.1. There are two ways to circumvent this problem. The first approach is to match applicants against a relatively small watchlist instead of the entire database. For example, the watchlist can contain several hundred persons of interest, for example people suspected to be actively involved in identity theft. In this case, a match threshold approach would work well. The second approach would be to use multiple biometrics. For example, if fingerprints were also enrolled, a front of house identification approach would be feasible, even for databases with millions of records (see Chap. 4).

10.1.1.2 Back of House

In general, when a new identity document is being requested, there is no need for it to be issued on the spot. Therefore, both of the methods presented in the previous section can also be implemented as back of house procedures, and the process is essentially the same. There are several advantages of this approach:

- The processing can be centralized. This equates to an easier and less expensive implementation, as it does not require software licenses and terminals at many locations throughout a country.
- There is a computational advantage in that the processing does not need to be conducted in real-time.
- Trained and dedicated investigators can review all cases. In this case, they have the time, expertise and resources to examine potential cases of fraud much more thoroughly than a front of house operator.

In addition to verification and identification, there are other uses for back of house biometric technology:

- Many databases have problems with data integrity, but this information is difficult to detect manually due to the sheer volume of data. Biometrics can be useful for data cleansing processes. For example, data entry errors can lead to the same person existing under multiple identities in a database. In essence, cases such as this have similar properties as cases of fraud, so similar techniques can be used to detect them. Similarly, experiments can be designed to uncover men accidentally labeled as women and vice versa, as well as other data integrity issues.
- Agencies that maintain large identity databases are often contacted by other agencies to help with their investigations. For example, consider a case where a person

has been found in a coma, and their identity is unknown. An off-line identification tool could be a service provided to other agencies to help conduct these investigations.

The primary disadvantage of back of house operations is that fraudsters are not "caught in the act", and a valuable opportunity for apprehension may be lost. However, as mentioned in the previous section, biometric decisions are rarely strong enough to justify immediate and decisive action anyway.

10.1.2 Estimating Levels of Fraud

Section 10.1.1 presented two ways in which biometrics can be used to help prevent the issuance of fraudulent identity documents. However, in most large scale identity databases, there will already be existing cases of fraud. Estimating the extent of fraud is difficult, and many organizations have no idea how common it actually is.

For our purposes, a case of fraud is defined as a single instance of a person obtaining an ID under another person's identity. Biometrics can be used to automatically detect this situation because there may be some match pairs labeled as genuine but are actually impostor (i.e. the application photo vs. the previous enrollment photo) and there may be some genuine matches labeled as impostor (i.e. in the case where the same person has committed fraud multiple times). In both cases, these match pairs will tend to exhibit themselves as outliers since they are drawn from one distribution, but labeled as the other.

There are a number of assumptions that are made with regards to the following analysis:

- People committing fraud do not actively change their appearance to a) look like their victim or b) look unlike themselves. If this is the case, it would alter the underlying match score distributions. However, the extent to which the match score distributions change would depend on the effort and skill of the fraudster, and is therefore very difficult to estimate. However, in our experience fraudsters make very little effort to alter their appearance beyond simple measures such as wearing glasses or changing their hairstyle, and a robust face matching algorithm already minimizes the impact of these factors.
- Mistakes made during the manual investigation process are not taken into account.
- The data integrity is assumed to be high. A ID labeling error in a database (i.e. accidentally assigning the wrong name to an identity document) has the same properties as a case of fraud.
- It is assumed that a fraudster has at least two enrollments in the database. One of these may be legitimate (i.e. their real driver's license or passport) or there may be a combination of legitimate and other false IDs.
- Genuine and impostor scores are independent and identically distributed, and covariance between the two is not modeled.

10.1.2.1 Running the Experiment *

The experiment is composed of two steps. In the first step, a large scale cross-match is conducted to find impostor matches with very high match scores, which are likely cases of fraud. In the second step, this list of candidates is culled by matching suspected fraudulent enrollments against other enrollments with the same name.

Step 1

The general idea of the first step is to observe random, high-scoring match pairs. Matches that score in this range are common for genuine matches, but very rare for true impostor matches. Therefore, this offers some evidence that a match is a possible case fraud.

It is generally not practical to conduct a full cross-match of a whole database if its size is in the millions, as this would lead to trillions of matches. Therefore, the database is randomly sampled to create a smaller, more manageable test set. Consider a database of enrollment images $\mathscr{D}$ with size $N = |\mathscr{D}|$. The process is as follows:

1. Randomly sample a proportion p of enrollments from $\mathscr{D}$. This will be the test set $\mathscr{T}$, and will be comprised of pN images. For example, if the database contains 1,000,000 images and a sampling rate of 10% (i.e. $p = 0.1$) is used, the test set $\mathscr{T}$ will contain 100,000 images.
2. Conduct a cross-match of all images in the test set, resulting in a set $\mathscr{M} = \mathscr{T} \times \mathscr{T}$ of matches. Exclude from $\mathscr{M}$ matches known to be between the same person. Therefore, in theory $\mathscr{M}$ only contains impostor scores. Since most matching algorithms are symmetric, it is actually only necessary to conduct half of the matches. In other words, if image A has already been matched against image B, it is not necessary to match B against A. The resulting number of match results is $|\mathscr{M}| \approx (pN)^2/2$. This is actually an upper bound, because the true number depends on how many genuine matches were excluded. However, the number of impostor matches generally far outweighs the number of genuine matches, so this is a good approximation. Using the previous example, the number of matches would be $|\mathscr{M}| \approx (0.1 \times 1{,}000{,}000)^2/2 = 5$ billion. This is a large number, but not unreasonable for today's technology.
3. Select a threshold t_1 to remove most members of $\mathscr{M}$, leaving only the highest scoring matches. In other words, the threshold is selected so that the false match rate is extremely low. For example, select t_1 such that the false match rate $\text{FMR}(t_1)=1.0 \times 10^{-6}$. In other words, only 1 in 10 million true impostor scores exceed the threshold.[1] Surprisingly, the correct match rate at this threshold may be relatively high. For example, it would not be unreasonable for a matching

[1] In order to select t_1 at this level, it is necessary to know the impostor score distribution with a high degree of precision. This should be established experimentally using a data set with the same properties as the one under investigation.

algorithm with an EER around 3%. to have a correct match rate of 70% (i.e. FNMR(t_1)=0.3) at a false match rate of 1.0×10^{-6}.

4. The resulting set is $\mathscr{I}$, the set of highest scoring matches: $\mathscr{I} = \{m \in M \mid m \geq t_1\}$. Assuming no fraud, the expected size of the set is $|\mathscr{I}| = |M| \times \mathrm{FMR}(t_1)$. In other words, this is the number of true impostor matches that would be expected to be above t_1. For the running example, the expected number of impostor matches in $\mathscr{I}$ is $5{,}000{,}000{,}000 \times (1.0 \times 10^{-6}) = 5000$.

Step 2

The second step is to add matches from the set $\mathscr{I}$ to a set $\mathscr{F}$ by excluding people who look similar to their previous enrollments. For each match $m(a_1, b_1) \in \mathscr{I}$, a_1 is an enrollment labeled as person a, b_1 is an enrollment labeled as person b, and $a \neq b$. We have one additional piece of information that can be used to test the likelihood that either a or b is a case of fraud. Namely, if either a_1 or b_1 is a case of fraud, they will have a low score when matched against other enrollments for a and b, respectively. We now select a second match threshold t_2 that is used to eliminate most impostors from $\mathscr{I}$.

For example, assume that $|\mathscr{I}| = 5500$, and contains 5000 actual impostor matches, and 500 actual cases of fraud. Further, assume t_2 is selected such that the false non-match rate is 10% (i.e. FNMR(t_2)=0.1). In other words, the probability that a true genuine match scores above t_2 is 90%. Consider the match $m(a_1, b_1) \in \mathscr{I}$ where both a and b have other enrollments in the database, labeled as a_0 and b_0 respectively. Let $m_a = m(a_0, a_1)$ and $m_b = m(b_0, b_1)$. If $m_a < t_2$ or $m_b < t_2$, add m to $\mathscr{F}$. In other words, if either a or b does not appear to match their other enrollment, add m to our set $\mathscr{F}$. To see the impact of this, we must look more closely at m. The match m will fall into one of two categories:

Non-fraud Assume that m is not a case of fraud. Therefore, a and b are two different people who just happened to look like each other (according to the match engine), and a and b should produce high match scores against their previous enrollments. In other words, m_a is genuine and m_b is genuine. The probability that $m_a \geq t_2$ and $m_b \geq t_2$ is $0.9 \times 0.9 = 0.81$, and m is not added to $\mathscr{F}$. Consequently, 81% of the non-fraud cases will not be added to $\mathscr{F}$.

Fraud Assume that m is a case of fraud. In this case, a and b are the same people, and a_1 is mislabeled, b_1 is mislabeled, or both are mislabeled. Therefore, at least one of m_a or m_b is an impostor. We are interested in the probability that $m_a \geq t_2$ and $m_b \geq t_2$, which would mean that a real case of fraud is being excluded from $\mathscr{F}$. This probability depends on the false match rate at t_2. A conservative example would be a FMR(t_2) of 0.01 at a FNMR(t_2) of 0.1. Therefore, for this example, the probability of one of m_a or m_b being an impostor and both scoring above t_2 is about 1%. If they are both impostors, the chances are even less of them both scoring above t_2.

Step 2 succeeds in excluding most non-fraud cases from $\mathscr{F}$, while adding almost all actual cases of fraud. With our running example, of the 5000 true impostors in $\mathscr{I}$, 4050 would be not be added to $\mathscr{F}$, while 495 of the 500 actual cases of fraud would be added to $\mathscr{F}$. In summary, from our original hypothetical database of 1,000,000 records, and the subsequent 5 billion matches, we are left with a set $\mathscr{F}$ of approximately 1000 match results, about half of which are likely to be cases of fraud.

10.1.2.2 Manual Investigation

The goal of manual investigation is separate the true impostors from the frauds in set $\mathscr{F}$. Some of the matches will be obvious impostors, and the reason why they achieved such a high score will be due to idiosyncrasies of the matching algorithm. These matches can be quickly discarded, as they are inconsequential. In other cases, it will be obvious that the images contain the same person, and there will be an obvious reason why. For example, it may be due to a data entry error. However, at the end of the day, the process is bound to be difficult and manually intensive. In our experience, with such a large set of data under investigation, every conceivable boundary condition is likely to occur. For example:

- In the case of identical twins, there will be two people who have an almost identical biometric sample (assuming facial photographs are being used), but it is not fraud. For men and unmarried women, this can often be resolved simply using surnames. However, in other cases, it may be more difficult.[2]
- There are many cases where people can legitimately have duplicate IDs with different names. For example, undercover operatives may have multiple identities. In cases such as this, care must be taken not to compromise the safety and security of the undercover operative. Providing investigators with a "black list" of people to exclude from fraud investigations is in itself a major risk. There is no easy way to deal with this situation. A good option is that cases such as these are not processed and stored as regular IDs, but are issued through a special procedure with its own independent, and secure, records.

A significant portion of $\mathscr{F}$ will be impostor matches containing people who simply look very similar, and cannot be distinguished conclusively by inspection alone. Before labeling a match a case of fraud, additional investigation must be conducted. For example, one may look at the supporting documentation that was provided at the time of application. This will often have strong clues (such as handwriting style, previous address information, etc.) as to the authenticity of the applications. Most

[2] We are aware of a case in which identical twin brothers had the same first and last name, and only differed by their middle initial. This caused so much confusion at the driver's license authority that investigators required the brothers to both show up at their offices in person at the same time in order to confirm that they were in fact two different people. A photograph of them together, as well as a signed statement, was recorded for future reference.

driver's license and passport authorities will already have a fraud department with the resources and skills to conduct this investigation.

10.1.2.3 Estimating Fraud Levels *

After the completion of the manual investigation, a set of samples is left with true cases of fraud. The final step is to use this information to infer underlying fraud rates. A high-level outline is presented below.

First the following question is addressed: given a random person who has two or more IDs with different names, what is the probability that they will be detected in the experiment? Let p be the sampling proportion, i.e. the proportion of the database $\mathscr{D}$ that was included in the test set $\mathscr{T}$. Assume that a person has two enrollments in the database under different names. The probability that both of these images will be included in the test set $\mathscr{T}$ is $p \times p$. We are now interested in the probability that this match will be included in the set $\mathscr{I}$, which is created by filtering $\mathscr{M}$ to retain only high scoring matches. The fraudulent match pair is drawn from the genuine distribution, so we are interested in the probability of a correct match, which is 1.0-FNMR(t_1). Finally, we need to know the probability that the match is added to the set $\mathscr{F}$. This is the probability that an impostor matched against the previous enrollment match scores below t_2, or 1.0-FMR(t_2). Thus, the final probability is as follows:

$$\text{probability of detection} = p \times p \times (1.0 - \text{FNMR}(t_1))(1.0 - \text{FMR}(t_2))$$

For the example of the previous sections, we have $p = 0.1$, FNMR(t_1) = 0.3, and FMR(t_2) = 0.01. Therefore, the chance of detecting a case where an individual has two images with different names in the same database is about 0.7%. Based on this value and the number of cases actually detected, one can estimate the actual level fraud. For example, if 5 true cases are detected, we can estimate that there are approximately $5/0.007 \approx 715$ cases in the full database.[3]

As it turns out, as people commit more fraud, the probability of their detection increases rapidly. This is beneficial, as many fraudsters are prolific. Once they learn the process to obtain fraudulent IDs, they are likely to do it many times. In fact, several cases in which people have obtained over 100 instances of fraudulent ID from a single issuance authority have been seen. The impact of this on detection probabilities is dramatic, because their chance of detection grows quadratically. This is illustrated in Fig. 10.2, which shows how the probability of detection increases with the level of fraud committed. For this example, people with more than 30 enrollments are almost certain to be detected, despite the fact that they are buried in a database with a million enrollments.

In summary, the proposed technique is not very useful as a method for exhaustively finding all existing cases of fraud in a database. In fact, it is unlikely to find

[3] The analysis is more complicated when the more general case is modeled where fraudsters have an arbitrary number of fraudulent IDs in the system.

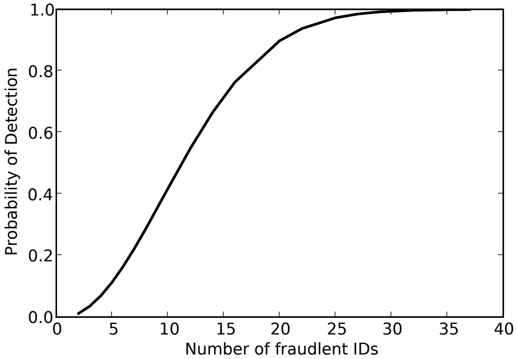

Fig. 10.2 The probability that a random fraudster is detected during an experiment, based on their number of fraudulent IDs. The graph is based on a hypothetical database and matching algorithm, using a sampling rate of 10%, an operating point t_1 with a correct match rate of 70%, and operating point t_2 with a correct non-match rate of 99%. The x-axis measures the number of fraudulent licenses held by the fraudster, and the y-axis is the probability of their detection. For example, for a person who holds 10 licenses, there is approximately a 40% chance that at least one fraudulent pair will be included in the experiment results.

any given person who has only two or three fake IDs. However, the probability of detection increases very quickly with the corresponding rate of fraud. Therefore, it can be used to find most of the habitual offenders. Furthermore, through careful reasoning, the likelihood that any given case of fraud will be detected can be determined. In turn, this can be used to estimate underlying fraud levels. However, it should be kept in mind that the procedure is based on a number of assumptions and estimations, so any reported outcomes have a high degree of statistical uncertainty.

10.2 Using Biometrics as Legal Evidence

In many criminal cases, in order to prove that a suspect is in some way associated with a crime, it is first necessary to prove that two images, video clips, or sound recordings contain samples from the same person. The proper use of biometrics in a legal settings is a broad subject, and could easily fill up a volume of its own. However, it is worth noting that many of the ideas and techniques presented in this book are fundamentally relevant.

Human fingerprint recognition for forensic purposes is well established, and has been used in legal proceedings for over a century [1]. More recently, there is some precedence for the use of forensic speaker identification [2]. Traditionally, the approach has been to rely on human *expert witnesses* who base a decision on their experience. For example, they may say something along the lines of “In all my 20 years in the field, I’ve never seen a match that so clearly ...”. This form of argument is known as an *appeal to authority*. The defense or prosecution is basically claiming that the witness is a recognized expert in the field, and therefore should be trusted. On one hand, the witness will often be correct. However, on the other hand the intuitions of an expert are not a suitable replacement for a rigorous, scientific, and probability-based argument.

Expert witness testimony is being increasingly challenged in courts [1]. People are inherently biased, often without realizing it. For example, expert witnesses may be subconsciously motivated to support a particular outcome, depending on who is paying for their time. Furthermore, they often have an implicit assumption of guilt or innocence, which will invariably impact the manner in which they read and interpret the available evidence.

10.2.1 Advantages of Biometrics

The field of biometrics offers an attractive alternative to expert witnesses for several reasons:

- A single piece of evidence is never conclusive - its purpose is to add weight to a given hypothesis or its alternative. In theory, a decision is reached "beyond reasonable doubt" by a jury who judiciously considers all evidence available. Therefore, each piece of evidence should have an associated probability value. Expert witnesses may provide a probability estimate, but these are rarely based on empirical studies, and tend to be justified with a great deal of hand waving. A primary advantage of biometrics is that the field is fundamentally grounded in statistical reasoning (see Chap. 7). Therefore, biometric "decisions" come with a probability estimate (based on the underlying genuine and impostor match score distributions) and a measure of statistical uncertainty (based on the size of the experiment).
- Biometric matching algorithms do not hold personal prejudices. This is not to say they are unbiased or that the results cannot be misinterpreted or misused, but there is a degree of subjectivity and impartiality that is difficult to achieve with human testimony.
- Biometric authentication offers a high degree of transparency, and is open to scrutiny by both defense and prosecution. Algorithms will inevitably have inherent biases due to their design and internal model. However, where these biases exist, they can be uncovered through a process of empirical evaluation. Once again, this is generally not possible with expert witnesses.
- Biometric algorithms have no memory. Therefore, it is possible to use them to test a series of different scenarios and hypotheses, with the outcome of one not influencing another.
- Biometric matching is very fast, allowing one to conduct large scale experiments that would not be feasible for human subjects. For example, to test an algorithm's ability to distinguish images of a particular population demographic, tests can easily be run that include thousands of individuals and millions of matches.

10.2.2 Disadvantages of Biometrics

Despite the advantages, it should be emphasized that computers are not infallible. People may be inclined to put blind faith in the output of an algorithm due to a feeling that computers are deterministic, and do not make mistakes. This is entirely wrong.

In fact, the title of this chapter, "Proof of Identity", is a misnomer because no proof is actually offered, at least in a *deductive* sense. In reality, biometric verification is based on probability distributions, and therefore the "proofs" are better thought of as *inductive*. In other words, a verification decision is based on reasoning along the lines of: "this is probably a genuine match because previous genuine matches I've seen tend to score in this range". Furthermore, the matching process itself is based on an underlying model, which makes assumptions of its own, and may be biased in unpredictable ways.

One of the reasons that courts are hesitant to accept the outcome of automated decision making processes is that people are uneasy with the idea that a legal outcome (which may ultimately be a person's freedom, or even life) may be falsely determined by a programming or processing error. This is a valid point and deserves careful consideration. However, it should also be kept in mind that human error has the potential to be just as costly, and is certainly not uncommon.

10.2.3 Match Score Distributions

One of the major themes of this book is that match score distributions lay at the heart of all biometric system. Correspondingly, almost all concepts in biometrics can be interpreted in terms of genuine and impostor score distributions. Another major theme is that these distributions do not exist independent of the context in which they are being used. In other words, a verification algorithm does not have a "true" error rate that is universally applicable across the board. An algorithm that works well in some circumstances will invariably work poorly in others. The most important factors are data quality and the user population. This has profound implications for the manner in which testing should be conducted when building a proof of identity argument:

Data quality The accuracy of a matching algorithm is heavily dependent on the quality of the biometric data. For instance, imagine a grainy CCTV image has been obtained from a crime scene, and face recognition is being used to help establish the identity of the suspect. In this case, genuine and impostor distributions (and subsequent error rates) derived from sharp, well-lit, full-frontal, high-resolution images are completely irrelevant. Therefore, every effort must be made to collect data of a similar quality as the samples under consideration.

User population The population that should be used for the tests is another important issue. In particular, consider the following question: should a random sample from the entire population be considered, or should one be more selective? The resolution recommend is that the genuine and impostor distributions be established separately. For genuine match scores, other individuals from the same demographic should be selected. This will answer the question "how well is the algorithm able to confirm the identity for people with similar physical characteristics as the suspect?". For instance, if the suspect is a young Caucasian male, the ability of the algorithm to verify the identity of an elderly Indian woman bears no consequence. For the impostor distribution, the population should be selected from the the full range of likely suspects. For example, consider a voice recording that is being used for identification. It will normally be obvious if the speaker is male or female. Assuming it is male, the impostor scores should be sampled from a variety of other males. In essence, this answers the question "how well is the algorithm able to distinguish this suspect from other *possible* suspects?". Including samples from the entire population makes the recognition task too easy, and in this setting one should always err on the side of caution.

Sample size is another important consideration. It is vital to be able to ensure statistical significance of the results, and this is done by including a sufficient number of samples in the test set. These topics are covered in detail in Sect. 7.3.

Another important consideration is the manner in which the results are presented to the jury. First and foremost, one should not attempt to present the matching process as a mysterious black box that never makes mistakes. One of the advantages of using biometrics is that, when used properly, it opens the lid of the black box and allows everyone to peer in. Most core concepts behind biometric matching can be explained intuitively to general audience (see Part I of this book), and every effort should be made to do so.

10.3 Conclusion

The focus of this chapter has been on using biometrics as a technique to help "prove" identity, in both business and legal contexts.

As seen in Sect. 10.1, biometric matching can provide a valuable contribution to improving POI procedures, as it provides a direct, ideally non-alterable, link to a person's physical identity. Having stronger identity management systems has significant benefits not only to government, but also to businesses and the wider community. With the threat and serious implications of identity crime in mind, biometrics should be considered a mandatory component of any future identity system.

Section 10.2 discussed the application of biometrics in a legal setting. It is not expected that the adoption of biometrics will occur overnight, nor is it likely that they will ever completely replace human testimony. However, we feel strongly that they have much to offer in the way of a scientific tool that helps establish identity. DNA evidence is used extensively court, and is often presented as the ultimate and infallible identifier. However, DNA matching is fundamentally based on probabilistic matching techniques. Furthermore, the matching process is largely automated and relies heavily on the use of sophisticated technology. In this sense, it mirrors the biometric matching process very closely. Therefore, there are no insurmountable barriers to the acceptance of biometrics, and considering the potential benefits, it seems likely that their use will gain momentum in the near future.

References

[1] Cole, S.A.: History of fingerprint pattern recognition. In: Ratha, N., Bolle, R. (eds.) Automatic Fingerprint Recognition Systems, pp. 1–25. Springer (2004)

[2] Rose, P.: Forensic Speaker Identification. CRC Press (2002)