Chapter 4
Multimodal Systems

Multimodal biometrics refers to the use of more than one source of information for biometric recognition [8, 9]. For example, a multimodal biometric system may use both iris recognition and fingerprint recognition to confirm the identity of a user. The use of multiple information sources helps to address some of the problems faced by real-world unimodal (also known as monomodal) systems, and multimodal biometrics will likely become increasingly common in future biometric deployments.

This chapter gives a brief, high-level introduction to the field of multimodal biometrics. The goals are to:

- Motivate the use of multimodal systems by outlining the primary advantages they offer over traditional systems (Sect. 4.1).
- Present the different approaches and modes of operation for multimodal systems (Sect. 4.2).
- Discuss information fusion and score combination (Sect. 4.3).
- Outline methods for the evaluation of multimodal systems (Sect. 4.4).

4.1 Advantages of Multimodal

The optimal biometric recognition system would be one having the properties of distinctiveness, universality, permanence, acceptability, collectability, and resistance to circumvention [9]. No existing biometric system simultaneously meets all of these requirements, however the use of more than one biometric can help lead to a system that is closer to these ideals. The advantages of multimodal systems stem from the fact that there are multiple sources of information. The most prominent implications of this are increased accuracy, fewer enrollment problems, and enhanced security.

4.1.1 Accuracy

The most immediate advantage of multimodal authentication is increased recognition accuracy. Multimodal systems fuse information for more than one source, each of which offers additional evidence about the authenticity of an identity claim. Therefore, one can have more confidence in the result. For example, consider two people who coincidentally have a similar facial appearance. In this case, there is a potential risk of a false accept for a system based purely on face recognition as a relatively high match score may be achieved when matching one against the other. However, if the same system also included fingerprint matching, it would be very unlikely that any given two people would have similar faces *and* similar fingerprint patterns. Therefore, the ability of the system to distinguish between people is increased significantly.

The previous example illustrated the use of multimodal matching for a verification system (one-to-one match). However, multimodal biometrics is particularly useful in an identification situation, as this involves many matches. Assume a fingerprint matching algorithm has a false match rate of 0.01%, which would be considered a high-accuracy system for fingerprints. With a database of 1,000,000 fingerprint images, one would expect approximately $0.0001 \times 1{,}000{,}000 = 100$ false matches for every identification query. In most cases this would be considered unacceptable performance, as it would be very laborious to manually examine all 100 false matches in the hopes of finding a correct one. Assume that an iris image is also stored for every person enrolled, and an iris matching algorithm is available that also has a false match rate of 0.01%. Under some reasonable assumptions, the probability of a fingerprint falsely matching an enrollment is independent of the probability that an iris will falsely match the same person, so the two probabilities can be treated as statistically independent. Therefore, the individual error rates can be multiplied together to find the expected error rate for the combined system.[1] The expected number of false matches for a multimodal identification query is $0.0001 \times 0.0001 \times 1,000,000 = 0.01$. This means that only one in every 100 queries would return a false match, which would be considered acceptable for most practical applications. This is an improvement of several orders of magnitude over using a fingerprint alone, and demonstrates the power of multimodal combination to boost the performance of identification systems.

4.1.2 Enrollment

Another advantage of multimodal systems is that they address the problem of non-universality, where a portion of a population has a biometric characteristic that is missing or not suitable for recognition. For many multimodal systems matching can be conducted even when one of the samples is unavailable or excessively poor

[1] This makes the assumption of an ‘and’ combination policy, rather than an ‘or’ policy.

quality. This will reduce the failure to enroll rate significantly. For example, consider the case of a person with damaged vocal cords who cannot speak, but would like to enroll in a system that uses voice authentication. In this case, their ability to enroll using an alternative biometric (such as a fingerprint) is a necessity.

Poor quality data is a common cause of enrollment errors, and ultimately false accepts and false rejects. The multimodal approach provides the system with a "second chance" to obtain or match a sample of sufficient quality, thereby increasing the robustness of the system.

4.1.3 Security

Multimodal systems have increased resistance to certain types of vulnerabilities, in particular spoof attacks. A spoof attack is where a person pretends to be another person by using falsified information. In the context of biometric systems, this involves creating and presenting an artificial representation of another person's biometric. For example, Japanese researchers have demonstrated how to create fake fingerprints using a commonly available material (gelatin) that has some success at fooling commercial fingerprint recognition systems [6]. The advantage of multimodal systems is that an attacker would have to be able to spoof two different biometric modalities simultaneously, which would be significantly more challenging.

4.2 Types of Multimodal Systems

There are several different ways that multimodal systems can be constructed, based on the sources of the biometric information and the way the system is designed. The term 'multimodal' sometimes refers specifically to the case where two or more different biometric modalities are in use (such as face and fingerprint), while the term *multi-biometrics* is more generic. Multi-biometric systems includes multimodal systems, as well as a number of different configurations.

Multimodal systems can be characterized by their sources of information or their organization.

4.2.1 Sources of Information

At a fundamental level, all multi-biometric systems collect and combine information from a variety of sources. However, the sources of the information differ from system to system. The following are the most common approaches:

Multiple modalities This refers to the situation where different biometric modalities are used, such as faces and fingerprints. The primary advantage of this approach is that it maximizes the independence between the samples. Therefore, a problem authenticating with one biometric is unlikely to impact authentication with the other.

Multiple characteristics This is the use of different instances (characteristics) of the same biometric. For example, one can match both thumbs for a fingerprint system, or the left and right eye for a retina system. In most cases, different biometric characteristics will have a high degree of independence. Furthermore, implementation is usually simple and cost effective because the same sensing equipment and matching algorithms can be used for each instance.

One characteristic, multiple sensors This uses multiple captures of the same biometric from different sensor types. For example, it may use both 2D and 3D face data. Another example is multi-spectral approaches, such as capturing a biometric with both the visible spectrum and infrared spectrum. The disadvantage of this approach is that if a biometric is not suitable for recognition (e.g. missing or damaged), the performance benefits of multiple captures will be minimal.

One sample, multiple algorithms This is the combination of multiple algorithms used to match the same sample. For example, different vendor face matching engines are applied to the same images. The advantage of this approach is where the algorithms have been developed independently, each will have its own strengths and weaknesses, and therefore may contain complementary information. However, since both algorithms are applied to the same data, there will be a degree of correlation between the results, and both algorithms will struggle with poor quality input.

Multiple impressions This technique uses multiple impressions of the same biometric characteristic. For example, multiple faces from a continuous video stream can be extracted and matched using a single matching engine. Another example is the integration of signals from multiple samples acquired at discontinuous intervals over an extended period of time.

Soft biometrics Multimodal systems can also use information from “soft” biometrics traits like height, weight, and eye color. These traits may have considerable variation in both the quality of the acquired data and temporal variation, so are of little value when used in isolation. However, some research has shown accuracy improvements when used in combination with other biometric traits [1, 3].

4.2.2 Modes of Operation

From the point of view of system flow, there are two common ways for a multimodal system to operate: in parallel, or in serial.

In serial mode, the various acquisitions and matching stages are conducted one after another. An example of this is a cascaded system, where if the user fails one biometric system, they use another biometric system, and the final output score is the fused scores of both. For example, the user first uses a fingerprint sensor, if this fails face recognition is used, and if this fails a hand print is required. The advantage of such a system is that many users will only need to use one sensor, streamlining the verification process. Some speaker verification systems use a variant of this where challenge response questions are asked until the user reaches a specified authorization level or the call is terminated.

For parallel systems everything is done somewhat simultaneously. For example, an ATM may require a person to have their fingerprint captured while they are looking at a camera for face recognition. There are security advantages to systems designed in this way because they are difficult to spoof. However, they may be more inconvenient and cumbersome to use from the user's point of view.

Combination Policy

There is a wide variety of ways that a multi-biometric system can be put together; either through logical combinations of match conditions or the application of various algorithms to combine scores. There is not one best combination technique for all circumstances, and deductions about what seems best from common sense may lead to poor outcomes in operation. The best advice is to choose a selection of potential combination techniques and run a series of scenario trials to determine which one will really perform the best for the desired outcomes (see Chap. 5).

4.3 Combination Techniques

Techniques for biometric fusion fall into three general categories, depending on the stage at which the combination is conducted: *feature fusion* combines low-level distinguishing features, *score fusion* makes use of multiple match scores, and *decision level fusion* logically combines accept/reject matching decisions.

4.3.1 Feature Level Fusion

As a general principle, better performance can be expected from feature level fusion than from score and decision level fusion. The reason for this is that the most information is available, which may be lost when fusion is conducted at higher levels. However, there are a few difficulties that make feature level fusion challenging and less common than other methods. First of all, the feature sets for the different information sources may be entirely different. For example, combining fingerprint and face features into a single model would not be a straightforward task as the matching algorithms (or distance metrics) used to compare these features are entirely different. Secondly, by simply concatenating features from different sources one would suffer from the problems associated with high-dimensional features spaces (known as the "curse of dimensionality"). Finally, for most commercial systems the feature data will be proprietary and not accessible to the system integrator.

One example of feature level fusion is the combination of fingerprint minutiae data from two different minutiae extraction algorithms. Both algorithms will have their own strengths and weaknesses, but when combined they will result in a more robust model of the fingerprint's minutiae. When this combined data is used for matching, better results are expected than when combining the final scores generated by the two algorithms because these are based on weak templates.

4.3.2 Score Level Fusion

Multimodal fusion is an active area of research, and the main emphasis tends to be on fusion at the match score level. There are several advantages to this approach:

- In general, score level fusion achieves better results than decision level fusion. The reason for this is that useful discriminative information (e.g. the confidence of a decision) is lost when the match score is disregarded.
- Match score fusion can be applied to most of the multi-biometric schemes discussed in Sect. 4.2. For example, it can be used to combine the scores of two fingerprint pairs, or the scores of a face match and a fingerprint match. Feature level fusion (discussed above) can be more difficult to apply in some circumstances.
- There is no need to have knowledge of the underlying algorithms or have access to feature information, as is the case for feature-level fusion. This is an advantage because it makes implementation easier. Furthermore, many commercial systems have proprietary formats for the storage of the feature data, so the information cannot be accessed directly.

There are two approaches to combination at the similarity score level: classification and score combination.

4.3.2.1 Classification

The idea of the classification approach is to consider verification as a classification problem with two classes: 'Accept' and 'Reject' [2]. Each authentication is represented by a feature vector that is composed of similarity scores from the various sub-systems. For example, assume face, fingerprint, and voice algorithms output similarity scores of 89, 76, and 58 respectively. This would lead to the feature vector [89, 76, 58], which would have an associated label of 'Accept' or 'Reject'. Training samples are collected, and machine learning or pattern recognition techniques (such as neural networks, support vector machines, decision trees, etc.) are used to build a model to distinguish the two classes. This classification model is applied to unseen data to make a verification decision.

4.3.2.2 Score Combination

Score combination involves taking several scores and applying a formula to combine them into a single score. Some examples include adding the scores together, taking the average, or selecting the minimum or maximum score. Several studies have concluded that the sum rule (a weighted average) is the best option due to its simplicity and high performance [4, 7].

One of the main issues that must be addressed for score level fusion is known as *normalization*.[2] The issue is that the similarity scores from different algorithms may not share the same underlying properties or score range. Therefore, simple approaches such as taking the average of two or more scores usually cannot be applied without first performing normalization. This can be illustrated with the following example. Consider Algorithm 1 that generates scores in the range [0...1] and Algorithm 2 that outputs scores in the range [0...100]. Obviously, taking a direct average will not work as the Algorithm 2 score will dominate the result. However, even when the score range is the same, the distribution of scores in this range (e.g. the mean and variance) may be entirely different. Score normalization is the process of transforming match scores from different sources into a standard distribution (both range and shape). Once this has been done, the individual scores can be compared against each other and combined accordingly.

When combining scores that have been generated from different biometric characteristics, the scores will have a high degree of independence. For example, if a person achieves a high fingerprint match score, this usually will not imply how well their face image will match. However, in some cases the match scores from different systems will be related to each other in some way. This will be particularly true when multiple algorithms are being applied to the same input data. This is illustrated in Fig. 4.1, which shows the results of applying two fingerprint matching algorithms to the same data. The effect of this is that the separation between the genuine and impostor scores is increased. A linear discriminant function is included that is able to distinguish between the two classes better than one based on the scores from only one algorithm (i.e. a horizontal or vertical line).

[2] Note that this is different from the concept of normalization for identification systems, which attempts to maximize the distance between vectors in the feature space (see Sect. 7.2.1.2).

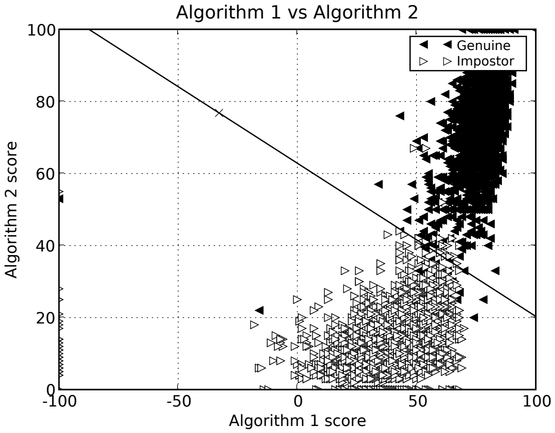

Fig. 4.1 Multi-algorithmic analysis of fingerprint matching results using a FVC 2002 data-set [5]. The plot is of a minutiae algorithm against a non-minutiae algorithm. The straight line through the points shows a separation between impostors and genuine points that is better than can be achieved using a single algorithm. The decision function for the line is that a match is genuine if 65*(Score$_{Algorithm2}$) + 17*(Score$_{Algorithm1}$) > 3752, leading to a FNMR of 1.6% and a FMR of 0.2%.

4.3.3 Decision Level Fusion

Decision level fusion is the highest level combination possible in the sense that all information about the matching process has been extracted except for a binary decision. In some circumstances, the raw match score is not available, such as for commercial systems that output "accept" or "reject". In this case, no information is known about the confidence of the decision. In other words, there is no way to distinguish between a strong match and a borderline match.

There is limited analysis that can be conducted for decision level fusion, and consequently the combination schemes are less sophisticated than for other modes of fusion. The most common approach is to use a majority voting scheme. For example, if there are three matching engines, the final decision is that which at least two engines agree on. When there is an even number of matching systems a voting scheme can lead to inconclusive results. More complex rules can be constructed using heuristics, such as "accept the user if any 6 out of the 10 fingerprint pairs match".

4.4 Evaluation

For feature and score level fusion, the evaluation of multimodal results is generally the same as the evaluation of other biometric systems (see Chaps. 7-9). The reason for this is that the output of multimodal systems is essentially the same as that from a unimodal system: a similarity score for verification systems (see Sect. 7.1), or a rank for identification systems (see Sect. 7.2). Therefore, the same modes of analysis can be conducted. In particular, false accept and false reject rates are used to report verification performance (e.g. ROC curves), and identification rates are used to quantify identification system performance (e.g. CMC curves). For decision level fusion match scores are not available, so performance rates are reported at a fixed operating point.

One of the primary advantages of multimodal systems is their ability to reduce enrollment errors. Therefore, an emphasis on failure to enroll rates should be an integral part of any multimodal evaluation.

4.5 Conclusion

This chapter has provided a high-level introduction to multimodal biometric solutions. The use of multi-biometric systems for enhancing system performance at both an algorithmic and system level is becoming increasingly important due to its numerous benefits over unimodal systems.

There are some disadvantages of multimodal systems, as they may be more expensive and complicated due to the requirement of additional hardware and matching algorithms, and there is a greater demand for computational power and storage. From a user's point of view, the systems may be more difficult to use, leading to longer enrollment and verification times. Furthermore, there are interoperability challenges related to the integration of products from different vendors. Despite these challenges, the field continues to be an active area of research because of the potential benefits of increased accuracy and security, and fewer enrollment failures.

As new biometric techniques are introduced they can often be combined with existing biometrics. Some recent examples include: iris and face, skin-texture and face, and skin pores and fingerprints. However, regardless of how systems are fused together, the final result is still a matching module that outputs a single score on which a decisions are made. Hence, the techniques described in Part II of the book apply equally to multi-biometrics systems as they do to unimodal systems.

References

[1] Ailisto, H., Vildjiounaite, E., Lindholm, M., Mäkelä, S.M., Peltola, J.: Soft biometrics-combining body weight and fat measurements with fingerprint biometrics. Pattern Recognition Letters **27**(5), 325–334 (2006)

[2] Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E.: Fusion of face and speech data for person identity verification. IEEE Trans. on Neural Networks **10**(5), 1065–1074 (1999)

[3] Jain, A.K., Dass, S.C., Nandakumar, K.: Soft biometric traits for personal recognition systems. In: Proc. of ICBA, pp. 731–738 (2004)

[4] Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell. **20**(3), 226–239 (1998). DOI http://dx.doi.org/10.1109/34.667881

[5] Maio, D., Maltoni, D., Cappelli, R., Wayman, J.L., Jain, A.K.: FVC2000: Fingerprint verification competition. IEEE Trans. Pattern Anal. Mach. Intell. **24**(3), 402–412 (2002)

[6] Matsumoto, T., Matsumoto, H., Yamada, K., Hoshino, S.: Impact of artificial gummy fingers on fingerprint systems. In: Proc. of the SPIE, Optical Security and Counterfeit Deterrence Techniques IV, vol. 4677 (2002)

[7] Ross, A., Jain, A.: Information fusion in biometrics. Pattern Recognition Letters **24**(13), 2115–2125 (2003)

[8] Ross, A., Jain, A.: Multimodal biometrics: an overview. In: Proc. of the 12th European Signal Processing Conference, pp. 1221–1224 (2004)

[9] Ross, A.A., Nandakumar, K., Jain, A.K.: Handbook of Multibiometrics. Springer (2006)