Technical Document

About FAR, FRR and EER

Version 1.0

About FAR, FRR and EER

Here, we discusses some general principles of biometric recognition systems, describes different classification errors and explains how the quality of two systems can be compared objectively.

Identification vs. Verification

A biometric recognition system can run in two different modes: identification or verification. Identification is the process of trying to find out a person's identity by examining a biometric pattern calculated from the person's biometric features.

In the identification case, the system is trained with the patterns of several persons. For each of the persons, a biometric template is calculated in this training stage. A pattern that is going to be identified is matched against every known template, yielding either a score or a distance describing the similarity between the pattern and the template. The system assigns the pattern to the person with the most similar biometric template. To prevent impostor patterns (in this case all patterns of persons not known by the system) from being correctly identified, the similarity has to exceed a certain level. If this level is not reached, the pattern is rejected.

In the verification case, a person's identity is claimed a priori. The pattern that is verified only is compared with the person's individual template. Similar to identification, it is checked whether the similarity between pattern and template is sufficient to provide access to the secured system or area.

Thresholding (False Acceptance / False Rejection)

We uses scores (also called weights) to express the similarity between a pattern and a biometric template. The higher the score is, the higher is the similarity between them. As described in the preceding section, access to the system is granted only, if the score for a trained person (identification) or the person that the pattern is verified against (verification) is higher than a certain threshold.

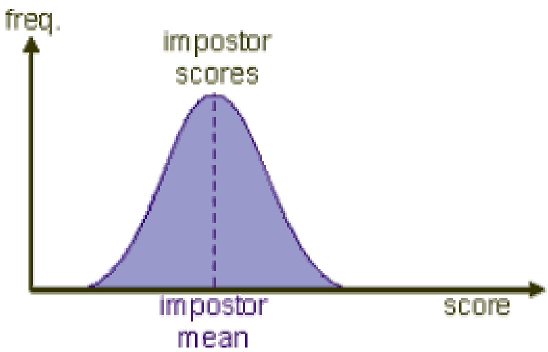
In theory, client scores (scores of patterns from persons known by the system) should always be higher than the scores of impostors. If this would be true, a single threshold, that separates the two groups of scores, could be used to differ between clients and impostors.

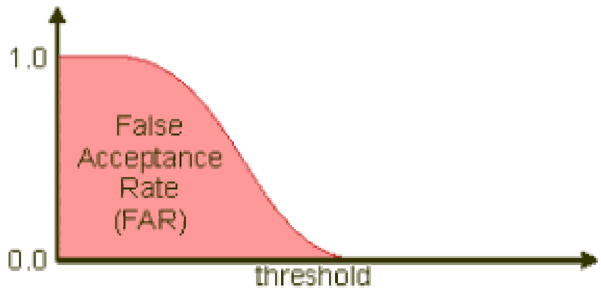
Due to several reasons, this assumption isn't true for real world biometric systems. In some cases impostor patterns generate scores that are higher than the scores of some client patterns. For that reason it is a fact, that however the classification threshold is chosen, some classification errors occur.

For example you can choose the threshold such high, that really no impostor scores will exceed this limit. As a result, no patterns are falsely accepted by the system. On the other hand the client patterns with scores lower than the highest impostor scores are falsely rejected. In opposition to this, you can choose the threshold such low, that no client patterns are falsely rejected. Then, on the other hand, some impostor patterns are falsely accepted. If you choose the threshold somewhere between those two points, both false rejections and false acceptances occur.

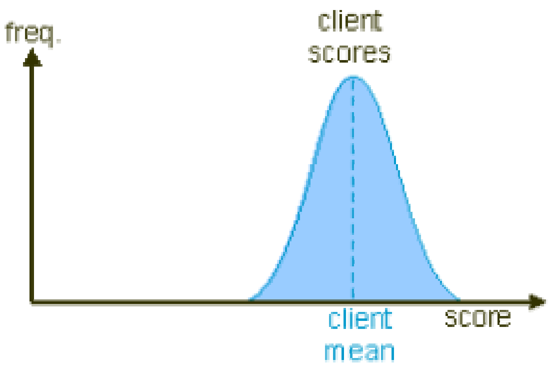
The following figures should help to achieve a better understanding of this topic. Think of a biometric verification system, which is tested with a large amount of test data. The test data consists of both impostor and client patterns. Let's first take a look at the impostor patterns. The belonging scores would be somehow distributed around a certain mean score. This is depicted in the first image on the left side. A gaussian normal distribution is chosen in this example.

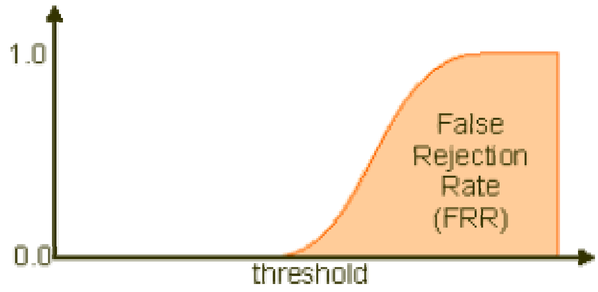
Depending on the choice of the classification threshold, between all and none of the impostor patterns are falsely accepted by the system. The threshold depending fraction of the falsely accepted patterns divided by the number of all impostor patterns is called **False Acceptance Rate (FAR)**. Its value is one, if all impostor patterns are falsely accepted and zero, if none of the impostor patterns is accepted. Look on the graphic on the right to see the values of the FAR for the score distribution of the left image for varying threshold.



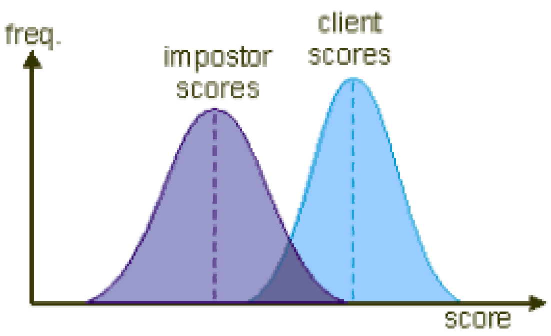


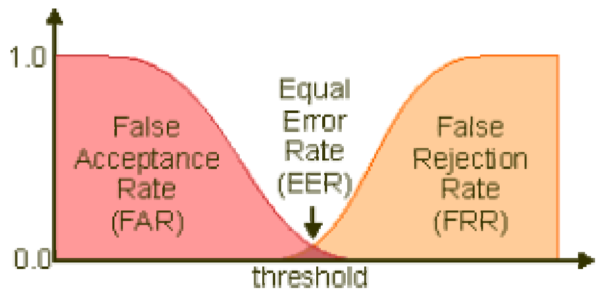
Now let's change to the client patterns. Similar to the impostor scores, the client pattern's scores vary around a certain mean value. The mean score of the client patterns is higher than the mean value of the impostor patterns, as shown in the left of the following two images. If a classification threshold that is too high is applied to the classification scores, some of the client patterns are falsely rejected. Depending on the value of the threshold, between none and all of the client patterns will be falsely rejected. The fraction of the number of rejected client patterns divided by the total number of client patterns is called **False Rejection Rate (FRR)**. According to the FAR, its value lies in between zero and one. The image on the right shows the FAR for a varying threshold for the score distribution shown in the image on the left.





The choice of the threshold value becomes a problem if the distributions of the client and the impostor scores overlap, as shown in the next image on the left. On the right, the corresponding false acceptance and false rejection rates are displayed.





Note that if the score distributions overlap, the FAR and FRR intersect at a certain point. The value of the FAR and the FRR at this point, which is of course the same for both of them, is called the **Equal Error Rate (EER)**.

Comparing biometric systems

Imagine the comparison of two biometric systems. The manufacturers of the systems just specify a single value for the FARs of them. Is this sufficient to compare both systems? The answer clearly is no, if the manufacturers do not provide the corresponding FRRs. In this case, it is possible that the system with the lower FAR has got an unacceptable high FRR.

But also when the values for FAR and FRR are given, there still exists the problem, that those values are threshold depending. Assuming that the threshold of the systems is adjustable, there is no reasonable way to decide if a system with a higher FAR and a lower FRR performs better than a system with a lower FAR and a higher FRR value.

The EER of a system can be used to give a threshold independent performance measure. The lower the EER is, the better is the system's performance, as the total error rate which is the sum of the FAR and the FRR at the point of the EER decreases. In theory this works fine, if the EER of the system is calculated using an infinite and representative test set, which of course is not possible under real world conditions. To get comparable results it is therefore necessary that the EERs that are compared are calculated on the same test data using the same test protocol.

Some effort is done to achieve large and publicly available test sets and protocols like the FERET and the XM2VTS databases for face recognition evaluation. But even if there exist some of those test sets for single biometric modalities, it still is very hard to compare systems relying on different ones (e.g. face and fingerprint).