

The Operational Context for Biometric Fusion

21 June 2006

Austin Hicklin,¹ Brad Ulery,¹ Craig Watson², Harold Korves¹

¹Mitretek Systems

²National Institute of Standards and Technology

Abstract

This paper is an overview of multi-biometric systems from an operational or systems engineering perspective. It discusses the purposes of multi-biometric systems in terms of accuracy, efficiency, robustness, applicability, and universality. Uses of multi-biometric data beyond the standard score- or decision-level fusion are discussed, including fusion-related techniques such selection and validation, as well as pre-match sample-level and template-level fusion. The paper discusses the architectures used in fusion using multiple matchers, and the implications for other issues such as data-sensitive matching, queuing issues, matcher decision implications, and human verification.

Contents

1	Introduction	3
2	Purposes for Multi-Biometric Systems	3
2.1	Accuracy	3
2.2	Efficiency	3
2.3	Robustness	3
2.4	Applicability	5
2.5	Universality.....	5
3	Multi-Biometric Data	5
3.1	Types of Multi-Biometric Data	5
3.2	Capture of Multi-Biometric Data	6
4	Uses for Multi-Biometric Data.....	6
4.1	Selection.....	7
4.2	Validation.....	7
4.3	Sample-level fusion.....	8
4.4	Template-level fusion	10
5	Multi-Matcher Fusion: Architectures and Implications.....	11
5.1	Parallel Fusion	12
5.2	Series Fusion	12
5.3	Data-Sensitive Matching	15
5.4	Fusion and Queuing Issues.....	15
5.5	Match Decision Implications	16
5.6	Human Verification	17
6	References.....	17

Acknowledgements

This work was performed for the National Institute of Standards and Technology, under contract through the U.S. Department of Interior, Contract NBCH-D-02-0039, Delivery Order D0200390057. Portions of this work draw on Mitretek Sponsored Research on Biometric Fusion conducted in 2004-2005.

We are grateful to our colleagues for their reviews and comments on earlier drafts of this document, including George Kiebusinski and Shahram Orandi (Mitretek).

This report was originally written as part of "Studies of Biometric Fusion" (NIST Interagency Report 7346), but separated and not published as part of that report to limit the scope of that work to empirical evaluation.

1 Introduction

This paper is an overview of multi-biometric systems from an operational or systems engineering perspective. It discusses the purposes for multi-biometric systems, the various ways that systems can take advantage of multiple types of data (which include but are not limited to fusion), and the system architectures used to combine multiple methods of processing.

2 Purposes for Multi-Biometric Systems

Although a number of biometric systems have been operational and successful for years, there is still room for improvement. Multi-biometric systems can — potentially — assist the designers and implementers of biometric systems in addressing a number of issues, including *accuracy*, *efficiency*, *robustness*, *applicability*, and *universality*.

2.1 Accuracy

Present-day biometric technology is imperfect, and leaves room for improvement in accuracy: TAR cannot be 100% while simultaneously keeping FAR to 0%. Fusion can improve accuracy by taking advantage of multiple sources of information to reduce the proportion of matching errors. For instance, a system that uses two fingerprints can correctly recognize more subjects (at a given FAR) than a system that uses only one. The additional information provided by the second fingerprint improves overall system accuracy, especially in cases where the first fingerprint is poor quality, distorted, or closely resembles that of another subject. Most discussions of fusion — and the remainder of papers in this series — focus on improving accuracy.

2.2 Efficiency

System efficiency (throughput, processing time, computational complexity, and financial cost) is of great concern to designers of most biometric systems, especially large-scale systems or systems that require very rapid responses. Efficiency can be improved by using multiple sources of biometric information and/or multiple types of matchers. For example:

- By using multiple matchers in series and allowing the earlier, faster matchers to make short-circuit decisions in the most obvious cases, the overall system can be made much faster than a single-matcher system, with limited loss of accuracy — if carefully designed and implemented.
- Somewhat counter-intuitively, access to more input data can make systems run both faster and more accurately. This is because a large amount of discriminating information can permit matching on fast or comparatively inexpensive matchers, when the equivalent accuracy with less data would require slower and computationally intensive algorithms.
- Some systems require human processing of cases that cannot be decided automatically, but human adjudication is expensive in both time and money. Use of specialty (“end-stage”) matchers to limit the number of cases that need to go to humans can reduce this expense.

2.3 Robustness

Robustness refers to the ability of a system to continue to function despite data problems or component level failures. Data problems such as poor sample (image) quality, database errors, and fraud are serious issues for most systems:

- In most real-world biometric systems, especially large and broadly distributed implementations, poor sample quality can be reduced and controlled, but it is unrealistic to assume that poor quality data can be completely eliminated.¹
- Biometric databases, like all databases, can contain errors (such as erroneous cross-references) that must be detected and resolved. These are described variously as database errors, data integrity errors, or metadata errors.
- Biometric fraud may take the form of evasion (in which the subject is avoiding identification) or spoofing (in which the subject is attempting to be falsely identified as a different subject).

The use of multiple biometrics increases system robustness and fault tolerance, by overcoming the deficiencies inherent in using single sources of data or single algorithms. Multiple biometrics can be used to lessen sensitivity to poor-quality data, erroneous input, or fraud. Many failures to enroll, false rejects, and false matches are due to characteristics of a specific sample: multiple biometrics can reduce the system's sensitivity to sample-specific noise. Since different matching algorithms have different strengths and weaknesses, the use of multiple algorithms is a form of risk reduction, limiting reliance on a single technology.

Some of the robustness of multi-biometric systems results from the increased accuracy of fusion; additional benefits accrue not through fusion but through data alternatives, redundancy, and validation:

- **Data alternatives** allow decisions to be made even when one biometric cannot be obtained, whether due to factors related to the subject or the system, or when one biometric cannot be processed, whether due to quality or component outages. For verification systems, data alternatives (biometric or not) permit legitimate users to use the system despite temporary problems such as bandaged fingers.
- **Data redundancy** lessens the risk of decisions based on a single sample, especially in cases of poor or marginal quality, database errors, or fraud. For example, if a set of fingerprints has the index and middle fingers swapped, the error might be detected with redundant data, such as a "slap" image in which the fingers were captured simultaneously.²
- **Data validation** is made possible by data redundancy, even if not used for fusion per se: should a reason exist to question a biometric identification, redundant information is critical because it can corroborate — or contradict — the determination. In cases of contested identifications, a mechanism needs to exist for independent verification that the biometric system did not fail, and that there were no administrative errors in the database. A system that cannot be verified cannot be proved right or wrong, which is a serious liability for an operational system; it is inappropriate to assume that a system that cannot detect such errors is necessarily error-free. Such verification can rely on additional biometric means, human validation, or non-biometric data.³ In our analyses here as well as in [FpVTE] and [SlapSeg],

¹ See [DataQuality] for a discussion of the causes and implications of poor data quality and data integrity errors in biometric systems.

² To give a simple non-biometric example, if a meeting is called for "Tuesday, 1 March 2006", the day of the week and the date are redundant. However, since 1 March 2006 is on a Wednesday, the redundant information permits the detection of the error, but not correction of the error. A third piece of redundant information (such as "tomorrow, Tuesday, 1 March 2006") may make it possible to correct the error without going back to the source.

³ A high-profile example of the value of validation is in forensics, where the availability of DNA has made it possible to validate (or invalidate) convictions, sometimes years after the fact. This is

we have found that multi-biometric data or multiple matchers can be used effectively to isolate data integrity errors. This is discussed further in [StudiesOfFusion, Section 2.4].

2.4 Applicability

Not all biometric modalities are appropriate for all uses. Systems that do not have to rely on any pre-existing data have much broader options than systems that require searches against legacy databases. Currently, multi-modal biometric systems that incorporate both fingerprint and face are applicable to the broadest variety of uses, but this is changing as other modalities (most notably iris) are coming into broader use. Note that these issues concern multi-modal systems, but not necessarily multi-modal fusion.

2.5 Universality

The use of multiple biometric instances (such as irises from different eyes) or biometric modalities (such as face, fingerprint, and iris) increases the universality of the application, so that subjects with unavailable or unusable characteristics can still be processed. Such unavailable characteristics include:

- Permanently missing characteristics such as amputations.
- Temporarily unavailable characteristics due to factors such as bandages.
- Obscured characteristics for social reasons due to factors such as veils or avoidance of contact.
- Intrinsically poor-quality characteristics.
- Subjects who have persistent sample acquisition problems.

3 Multi-Biometric Data

3.1 Types of Multi-Biometric Data

The types of biometric data that can be used in multi-biometric systems include:

- **Samples:** multiple samples acquired from the same source, such as multiple images of a single fingerprint, images of the same face, or recordings of a speaker. Some sources (e.g. [SC37-24722]) further divide multi-sample into multi-sensorial and multi-presentation, which differ based on whether substantively different types of capture devices (sensors) were used. Note that a multi-sample system may involve collecting multiple probe samples, or may involve retaining multiple samples in the gallery, such as when an additional enrollment is added for every encounter with a subject.
- **Instances:** multiple instances of the same biometric modality, such as fingerprints from multiple fingers, or images of both irises.
- **Modalities:** multiple biometric modalities such as a combination of a subject's fingerprints, face, irises, and voice.
- **Metadata:** information related to the biometric data or the subject, such as measures of sample quality, location and date of sample collection, or demographic ("soft biometric") information such as gender, height, or age.

due not just to the accuracy of DNA, but also to the value of DNA and fingerprints as validation methods for each other, especially in the troubling cases of administrative error or fraud.

As a side note, multi-sample data can be created from a single sample through various transformations, to account for possible distortion or translation issues. Such faux multi-sample data can then be used in fusion.⁴

3.2 Capture of Multi-Biometric Data

Multi-biometric data can be captured in a variety of ways, depending on system requirements.

- **Collection at a single encounter** — all of the desired samples, instances, and modalities are captured from the subject at one time, though not usually simultaneously. True simultaneous capture is possible in some cases, such as cameras that capture both irises and the face at once, or slap fingerprint scanners that capture four fingerprints in one image.
- **Collection in series** — either multiple collections before the first attempt to match, or a series of collection-match events:
 - **Rejection and recapture** — if the initial sample is of inadequate quality (or fails to match), the sample is recaptured.
 - **Fail-safe collection** — for a verification system, if the initial attempt to match is unsuccessful, a different type of authentication (biometric or not) can be used.
 - **Secondary processing** — for a watchlist system, subjects provisionally identified may have additional data collected for a definitive identification or a search against additional systems.
 - **Successive collections** — for a verification/access control system, successive access points (e.g. doorways) can each have a collection device and matcher, so that full access requires a successful match in every case.
- **Challenge-response collection** — for some verification systems, a limited additional layer of security against fraud can be added by enrolling a variety of multi-biometric data (such as all ten fingerprints) and having the system randomly select the instance to be presented.
- **Gallery-side fusion** — enrollment of samples from multiple encounters with the subject enables the use of fusion without multi-biometric data capture. While the collection of multiple probe samples increases both complexity and time for collection, the retention and use of multiple gallery samples is a database issue that is invisible to subjects. Some identification systems enroll all probes, even if they were successfully matched, increasing accuracy for future searches with marginal quality or distortion issues.

4 Uses for Multi-Biometric Data

What systems do with multi-biometric data varies dramatically. The later papers in this collection focus on score-level and decision-level fusion, which are post-match transformations of the results of comparisons. This section addresses other uses of multi-biometric data:

- **Selection** — the use of only a subset of the data collected.

⁴ Some algorithms attempt to normalize input samples before processing to minimize distortion or translation issues. For example, face systems can attempt to correct pose, and fingerprint systems can attempt to account for uneven pressure, by manipulating the input images. Since detecting distortion is often imperfect, some systems apply a variety of transformations and then use these multiple manipulated samples as inputs, fusing the results. This is an interesting case of fusion using a single sample and single matcher, but with multiple methods of preprocessing. Other examples include multiple searches of a single face image with differing exposure or color balance. Similar transformations can be applied to templates rather than samples.

- *Validation* — the use of some of the data to check the integrity of the other data.
- *Pre-match fusion* — the use of transformations of the data used for comparison before matching is performed:
 - *Sample-level fusion* — the combination of multiple samples to form a single sample.
 - *Template-level fusion* — the combination of multiple templates to form a single templates.

4.1 Selection

Selection methods all deal with the selection of the best (usually highest quality) of a set of data. Most of these require effective automated quality metrics. Selection can take place at collection, for processing/matching, or at enrollment:

- *Selection at collection* — One or more samples can be acquired per instance/modality, with only one retained for further processing.
 - *Rejection and recapture* — The most trivial case is a system that only retains a single sample, but if the initial acquisition is inadequate (poor quality), multiple samples are acquired until one is acceptable or (in some cases) a given number of attempts is reached. If the rejection determination is not made while the subject is still present, recapture is problematic or impossible.
 - *Autocapture* — Similar to rejection and recapture, but multiple samples are captured for every subject, and the highest quality sample is selected automatically. This is preferable to rejection and recapture because improved quality can be realized even if multiple samples exceed the rejection threshold, or if none do.
- *Selection for processing* — When multiple instances are available, a system may be designed not to use all of them for matching, due to computational cost or diminishing returns. For example, some AFISs use all ten fingers for filtering, but only use two fingers in the main match stage, for example, using the index fingers if they are of acceptable quality, or the two highest-quality fingers otherwise.
- *Selection for enrollment* — Multiple samples per instance/modality are available, but only one is retained in the gallery for future searches. This is most frequently encountered after a successful search, when there are several options: 1) leave the original sample in the gallery, 2) compare the quality of the probe and gallery samples and retain the better sample, 3) keep the more recent sample in the gallery to minimize data aging (which usually means retaining the probe sample), or 4) retain both in the gallery (sometimes known as gallery-side fusion). Many systems have implemented means of using quality metrics to select the better sample. [Uludaga-04] describes a means of using the data content in multiple templates to select the template most representative of the available set of templates.

Selection is unlikely to be optimal. If the discarded samples/instances are truly useless or have no independent data content, this approach is appropriate. However, if usable inputs are captured but ignored, this runs counter to the concept of fusion: discarding data is rarely optimal in terms of accuracy.

4.2 Validation

Validation methods use some of the data as a way of checking data integrity (as was discussed above in Section 2.3). Two examples are the traditional use of rolled and slap fingerprints for cross-validation, and the use of additional data for human review.

- Traditionally, full sets of both rolled and slap fingerprints were taken for criminal arrest or background check transactions.⁵ The rolled fingerprints were used for matching, while the slaps were used to verify that the rolls were in the correct sequence. Without the redundant data, data integrity errors would pass into the system unnoticed. Such checks were once performed manually, but now automated sequence checks are built into many livescan devices.
- Some systems collect fingerprints and face images, but do not use the face images for automated matching, only using the face for human review of problem cases, such as egregiously poor-quality samples or potential data integrity errors.

From a systems perspective, the cross-checking of rolls and slaps and the human review of problem cases are processes that take advantage of additional data to improve system accuracy and robustness. Such processes are not fusion per se — but the inputs and purposes remain the same as for fusion.

4.3 Sample-Level Fusion

Traditionally, levels of fusion are defined as sample-, template-, score-, and decision-level fusion. These levels commingle two very different processes: sample- and template-level fusion are pre-match transformations of the data used for comparison, while score- and decision-level fusion are post-match transformations of the results of comparisons.

Sample-level fusion is the use of multiple samples (e.g. images) to create a single sample. Sample fusion is used in some contexts to create a new sample that is composed of the union of the areas of the constituent samples. For example, a rolled fingerprint image from a livescan device is actually a composite of a series of images (usually one or two dozen) collected as the finger rolled across the platen; the device combines all of the images to create one composite image, larger than any of the individual images, and (if done properly) limiting localized blurring or distortion in the overlapping regions. [Jain-02b] discusses “fingerprint mosaicking” a similar process that uses multiple images from small sensors to create a composite image. Sample-level fusion can also be used by 3D face systems that work from images collected by multiple cameras.

Sample-level fusion can create artifacts, such as shown in Figures 1 and 2. While subtle artifacts such as these may not be serious in some uses, they are of concern for forensic use, especially if the unfused samples are not retained.

⁵ Many background check applications are in the process of moving to slap-only transactions.



Figure 1: Example of an artifact created in sample-level fusion. The Z-shaped crossover (highlighted) does not exist in the actual fingerprint. The distorted/smeared areas are due to the “stitching” methods used to combine images that do not correspond precisely.

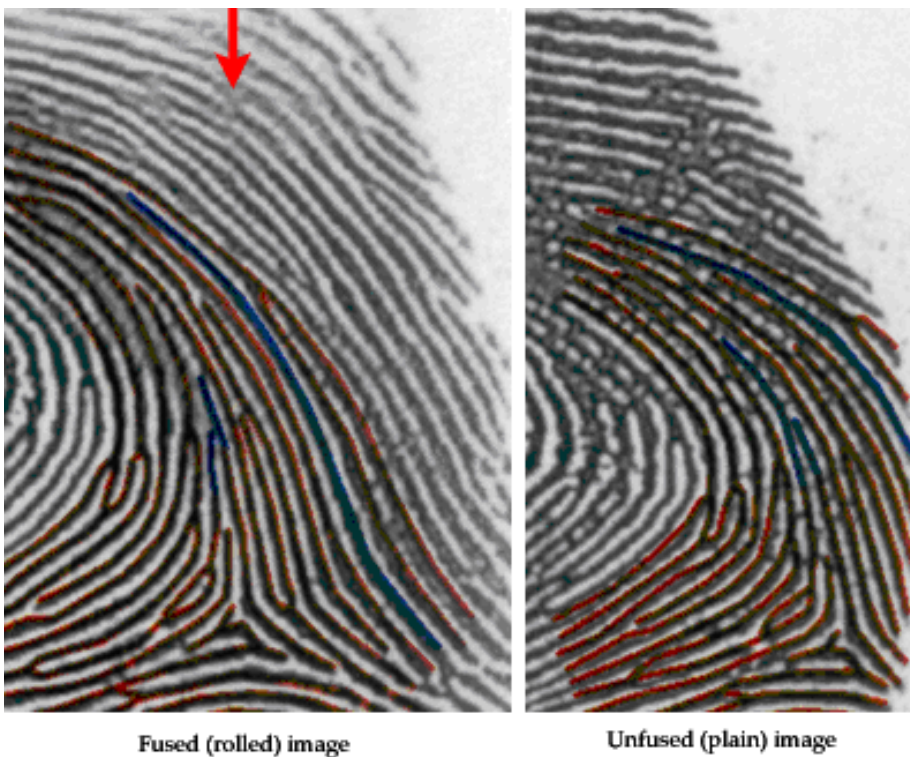


Figure 2: Ridge structures created in sample-level fusion. The structure of ridges (between the blue highlights) differs between the fused “rolled” image on the left and the plain image on the right. Note that there are linear discontinuities below the arrow that indicate stitching errors. [Senese-06]

4.4 Template-Level Fusion

Template-level fusion is the use of multiple templates (feature sets) to create a single template. Most current uses of template-level fusion are based on templates from the same biometric instance, acquired from multiple feature extraction algorithms or multiple samples; these can be straightforward to implement because of compatible templates and features. Dissimilar templates (from different modalities or instances) can be fused, but this is a research area that hasn't yet been shown to be effective, e.g. [Ross-05].

Multi-extractor and multi-sample template-level fusion are implemented in a variety of operational biometric systems. The primary benefit is that the resulting template is more robust than a template created using a single sample and single feature extraction algorithm: the presence and characteristics of each feature can be corroborated and quantified in terms of feature-specific quality values. Such feature-specific quality values are used effectively in some matchers.

- *Multi-extractor template-level fusion* is the use of different feature extraction algorithms are used to create templates for a single sample. For example, the FBI's IAFIS combines the results from multiple algorithms to determine each fingerprint's pattern classification, as well as using multiple algorithms to detect minutiae. This method can be used to mitigate any limitations of a single feature extractor, and can be used to determine feature-specific robustness or quality, based on whether the algorithms concur on the feature's presence and characteristics.⁶ Figure 3 shows an example of variation between different feature extractors: the weighting of individual features in the matching process can be based on the degree of correspondence for each feature as defined in multiple templates.

⁶ In an analogous manual process, the feature sets (templates) for latent fingerprints are usually created by human fingerprint examiners, with assistance from automated tools. When a latent fingerprint from a particularly important case is searched, the features may be defined by several examiners, either preparing separate searches and/or working together to create an optimal feature set.

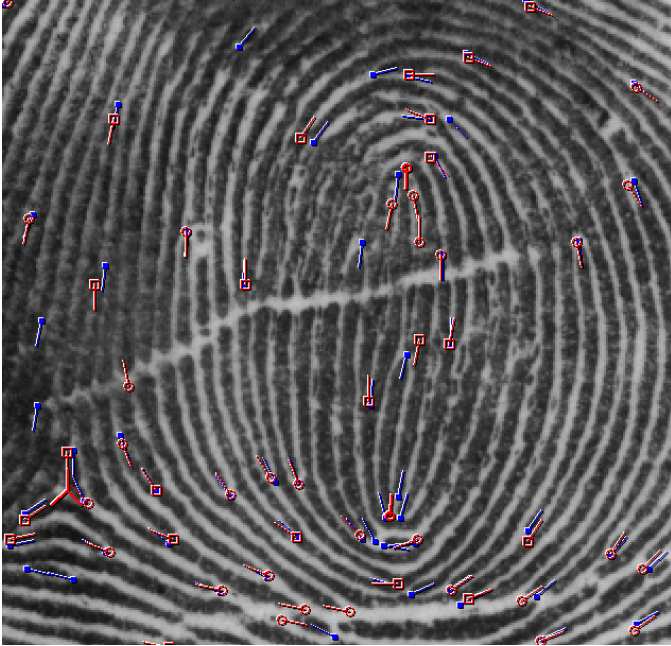


Figure 3: Comparison of templates generated by different feature extractors

- *Multi-sample template-level fusion* is the use of a single feature extraction algorithm is used to create templates for multiple samples of the same biometric instance. For example, given multiple fingerprints from the same finger, a template is created for each image, then all of the templates are combined to create a single composite template. The resulting template will cover an area larger than any of the individual templates, but can account for distortion, and can measure the quality (repeatability) of each individual minutia. In some cases, a mix of template and image fusion can be effective. For example, an algorithm that fuses a set of plain fingerprint impressions to create a single rolled impression can calculate the necessary deformations by fusing the templates, and then use that information to fuse the images.

5 Multi-Matcher Fusion: Architectures and Implications

When multi-matcher fusion is implemented, especially in large-scale systems, the architecture of the matcher subsystem must be considered.

A biometric matcher may just be a single algorithm that measures similarity (or difference) between biometric templates. However, many systems are more complex, using multiple algorithms to increase accuracy and/or throughput: most large-scale AFISs use two to five matching algorithms in various configurations. It should be noted that some systems are the beneficiaries of multi-matcher fusion even if the users of the system are not aware of it: a matcher component, if obtained without insight into its internal composition, may or may not itself be composed of multiple algorithms.

For large-scale systems, multi-matcher fusion can be classified by the system architectures used in implementation. Various fusion architectures have been implemented over the years, and are discussed and categorized in [Jain-00] and [Maltoni-03].

At the simplest level, multi-matcher fusion can be implemented using parallel or series architectures.

5.1 Parallel Fusion

In a parallel architecture, each of the matchers is used for all comparisons, and the results from one matcher do not affect how (or whether) matching is performed using the other matcher(s). Note that the matcher stages are “parallel” in the sense that they perform independent tasks, and are not necessarily performed simultaneously. Fusion in parallel can be optimal in terms of accuracy, but since it requires generating all scores in all cases, it can be computationally intensive, increasing cost or degrading throughput. Most discussions of multi-algorithm fusion implicitly assume that fusion is implemented in parallel. A simple example of parallel fusion architecture is shown in Figure 4.

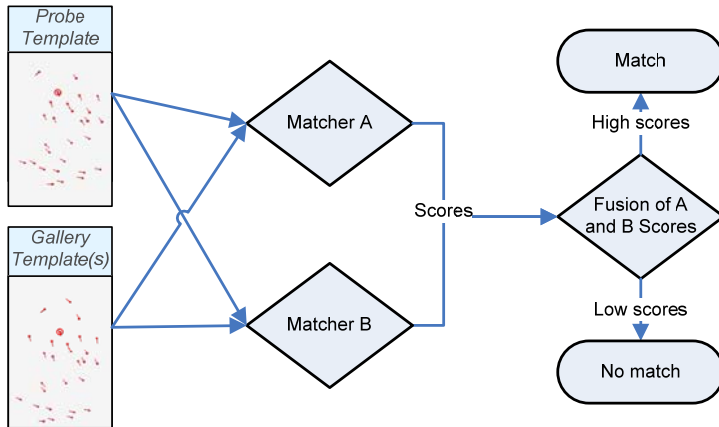


Figure 4: Example of fusion in parallel. Both Matcher A and Matcher B are used in all cases.

5.2 Series Fusion

In a series (or multi-stage) architecture, one matcher filters out some portion of the most obvious non-matches and/or matches, and subsequent matchers only work on the remaining indeterminate comparisons. An example of series fusion architecture is shown in Figure 5.

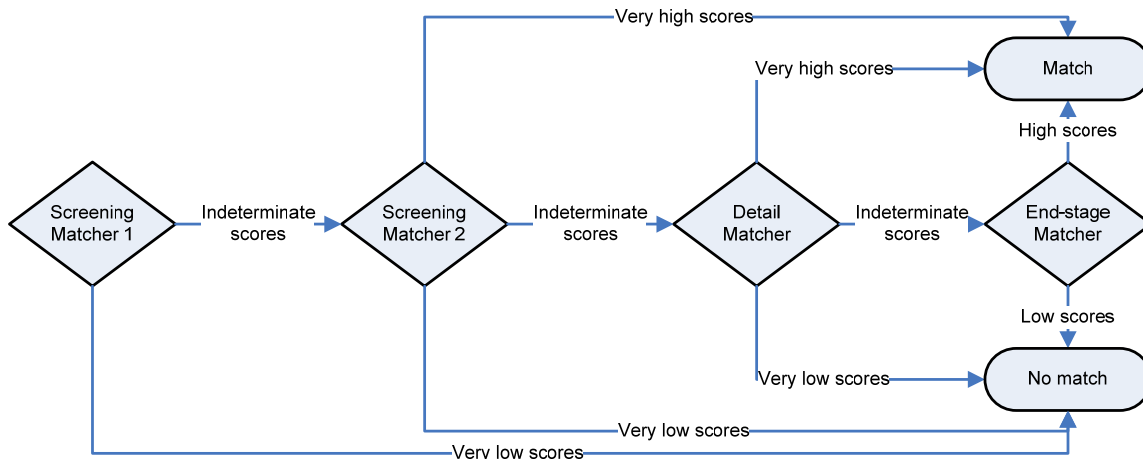


Figure 5: Example of fusion in series. Note that each matcher is only used if the previous matcher (s) return indeterminate results.

Fusion in series incorporates aspects of both unfused decisions and score-level fusion. Typically, the early-stage matchers (known as screening matchers) are faster, while the later-stage matchers are slower but more accurate. The screening matchers make unfused decisions in those cases where the first result is definitive: these are known as short-circuit decisions. The results from the later matchers can be fused at the score level with the results from the earlier stages; however, if the scores from earlier stages are not retained, the results are implicitly fused at the decision level. If the matcher algorithms are relatively independent, such decision-level fusion is unlikely to be optimal in terms of accuracy.

The matchers in multi-stage systems fill different roles than do single-stage matchers:

- Screening (or filtering) matchers are designed to rapidly exclude obvious non-matches, reducing substantially the number of comparisons required for processing by later matchers. Screening matchers make final non-match determinations, but in most implementations do not make any final match decisions: all potential matches are passed on to later stages. If the criteria for exclusion used in screening can be defined in terms of logical or database operations, the screening stage is sometimes known as a binning or partitioning process.
- Detail matchers, or matchers designed to be used after screening matchers, can focus on accuracy with less regard for computational complexity or throughput. The detail stage may actually consist of multiple matching algorithms fused in parallel.
- End-stage (or post-processing) matchers are detail matchers that are used only a relative handful of cases in which the primary detail matcher is unable to make a final match/non-match decision. This role is currently filled on many systems by human review, which from a systems standpoint can be viewed as a matcher stage. Because human review is slow and expensive, automated end-stage matching is an area of interest.

Matchers in series provide a means of tuning the tradeoffs between efficiency and accuracy:

- If the thresholds for the screening matchers are set to be very conservative so that relatively few comparisons are rejected at the first stage, there is a gain in efficiency with little degradation in accuracy.
- If the thresholds for the screening matchers are set to be aggressive so that many comparisons are rejected at the first stage, there can be a substantial gain in efficiency, but accuracy will suffer because of an increased proportion of false rejects.

Note that the optimal tradeoff between efficiency and accuracy depends on system-specific requirements. Use of short-circuit decisions and well-understood decision thresholds can translate into substantial improvements in efficiency.

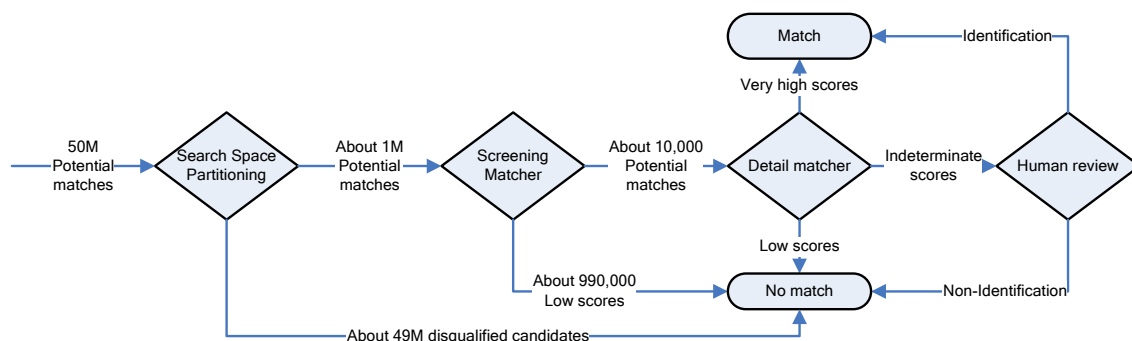


Figure 6: Multi-stage matching in the IAFIS ten-print system

The efficiency gain of multi-stage matching is shown in Figure 6, using the IAFIS ten-print system as an example. Most large-scale AFISs have similar architectures. Note that IAFIS uses two screening stages, a detail matcher, and human review, all fused at the decision level; the IAFIS latent matching system uses a similar architecture but for the detail stage uses two algorithms in parallel, fused at the score level. The early stages are very fast but cannot accurately make final match determinations, while the detail matcher is too computationally intensive to be used on all potential matches, so multiple stages provide a compromise.

Multi-stage matching can also be used based on multiple instances and modalities, as shown in Figure 7.

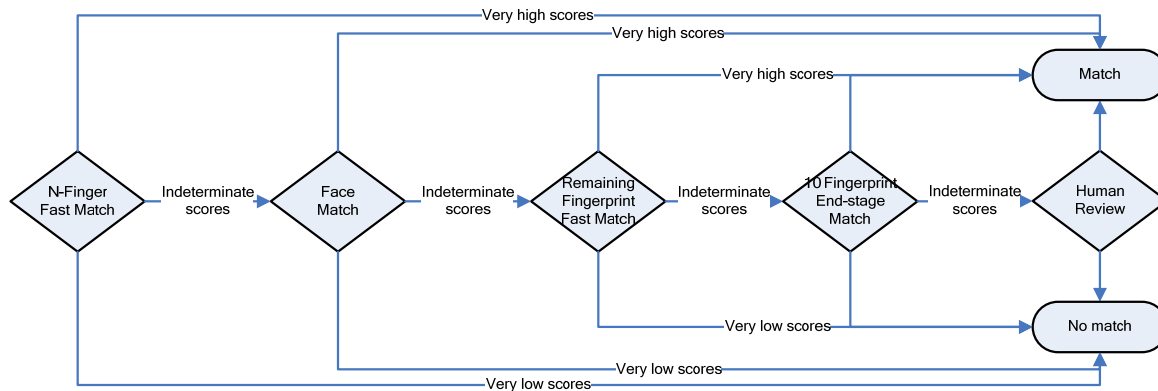


Figure 7: Example of multi-stage system that uses multiple instances and modalities. Each of the later stages can be fused with previous results at the score level; otherwise, the system is implicitly fused at the decision level.

The use of multi-stage matching requires consideration of a variety of issues:

- Multi-stage matching is most appropriate for those matchers that require computationally intensive processing, which is particularly true for fingerprints. Some other biometric modalities have an advantage in that detail matching is faster, such as the Daugman iris algorithms. Systems that can provide both high accuracy and high throughput in a single algorithm have less need for multi-stage matching to improve efficiency. However, addition of an end-stage matcher is still likely to improve accuracy, by serving as a means of arbitrating any indeterminate results from the original matcher.
- Screening matchers are designed or tuned for different operating points than single-stage or end-stage matchers. A screening matcher that is designed to eliminate 90% to 99% of the most obvious non-matches will use operating points corresponding to 1:1 FARs between 1% and 10%, which are far larger than the false accept rates necessary for detail matchers. Accuracy for screening matchers is driven by extremely low false reject rates for those operating points.
- At least in theory, a multi-stage system will never be as accurate as a system that fuses the results from all of the matchers in every case: short-circuit architectures are appropriate when it is impractical to use detail or end-stage matchers for every comparison.
- Because screening matchers reject such large proportions of potential matches, the false reject rates of screening matchers are likely to dominate the overall system false reject rate: care should be taken not to use screening matchers more aggressively than is necessary.

- Screening matchers eliminate most of the easy imposter comparisons performed by the overall system. Unless the stages are completely independent, the later stages will therefore have much lower accuracy than they would have against the overall population. Accurate evaluation of the performance of an individual later-stage matcher is complex and may be subject to misinterpretation.

5.3 Data-Sensitive Matching

Metadata characteristics of each sample can be used to determine how matching will be conducted, or which matching algorithms will be used. For such data-sensitive matching, the system architecture may be a hierarchy, or may contain multiple discrete paths.

Data-sensitive matching may be used based on characteristics that make the standard methods of matching less effective, such as poor sample quality or very old or young subjects. In addition, data-sensitive matching may be based on the relative importance of matches against specific subjects in the database, such as individuals of particular interest in a watchlist; such cases may have different accuracy requirements due to the increased value of a match/cost of a false non-match.

There are several ways that matching can progress differently based on the metadata associated with a specific subject:

- Different matcher score thresholds can be used in decision points. [Wein05]
- Scores can be normalized or fused differently based on metadata.
- Special preprocessing can be performed on the sample, such as “age progression” of a past face or fingerprint image from a child.
- Entirely different matching algorithms can be used.
- Human special processing may be used based on subject-specific triggers.

In some cases, partitions of the gallery may be created for certain subsets of gallery subjects. For example, separate galleries may be created to house poor-quality samples, or subjects of interest.

5.4 Fusion and Queuing Issues

As discussed previously, multi-biometric systems can be used to improve accuracy, efficiency, robustness, and fault tolerance. Usually, increased accuracy receives the greatest emphasis, as is the focus of most of our analyses. However, accuracy is only one criterion associated with the design, development, and operation of a large-scale identification system. If history is a guide, there are likely to be significant problems and concerns associated with issues such as system integration, manual processes, exception handling, and even customer service. Because fusion parameters are a key means of tuning system performance, there is a (somewhat surprising) relationship between fusion and these other system issues.

A large-scale identification system must address a variety of issues such as the relative costs and risks associated with identification errors, staffing, hardware limitations, acceptable wait time for subjects, and system capacity. Use of a model that considers biometric system accuracy in terms of network queuing theory can help to address such operational concerns. [Korves-05]

Some implementations of fusion permit tradeoffs between accuracy and efficiency that can be managed dynamically so that higher accuracy can be achieved during periods when high throughput is not necessary. During periods of low usage, a multi-stage identification system could be tuned so that a larger proportion of candidates are sent to slower but more accurate

matchers. During brief peak periods, slower processes can accumulate queues that can be eliminated as the peak demand diminishes. During longer peak periods the length of queues would have to be limited; this can be accomplished by tuning the system thresholds to exclude a larger proportion of matches from the slower matchers, with the side effect of lower overall accuracy. If a situation requires higher security for a certain period of time, overall accuracy can be increased by increasing the acceptable queue length. Note that some systems can have multiple queues, such as subjects waiting for initial processing, subjects waiting for secondary processing, and candidate matches waiting for human review.

Two constraints on such a system are the maximum acceptable queue length(s) and the minimum acceptable accuracy, defined in terms of FRR and FAR. Throughput is increased at the cost of increasing false rejects. Since candidate matches still would progress through the subsequent matchers, the false accept rate could be kept approximately constant.

Overall system capacity can be defined as the level of extreme system demand where the constraints of maximum acceptable queue length and minimum acceptable accuracy can no longer both be met.

5.5 Match Decision Implications

The fundamental purpose for biometric systems is, obviously, to make match determinations. However, the implications of these determinations vary substantially. These implications can be considered in two ways:

- **Importance** — The relative importance of errors (false accept, false reject, or failure to acquire/enroll) can be viewed in terms of dollar cost, throughput cost, inconvenience, possible threat, or legal ramifications.
- **Finality** — Relatively few systems make final decisions for both match and non-match determinations; most systems have some fallback procedures to provide final verification or to handle claims of error. A system that does not make a final and definitive determination is in effect a matching stage in a larger system — a filter before another more rigorous matching process such as a further biometric collection and matching, human verification, end-stage/specialty matching, or alternate non-biometric procedures. Determining the finality of a system's match determinations must be made in light of the system's error rates, prior probabilities, and importance of decisions.

Match decision implications are related to multi-matcher fusion because any system that does not make a final determination is in essence a system fused in series; any system analysis or measurement of accuracy must take that into account. If the systems are implemented in ignorance of each other, they end up implementing decision-level fusion, which may be warranted due to system requirements, but is unlikely to be optimal in terms of accuracy.

Negative ID systems generally require verification or secondary processing for matches (to eliminate or limit false matches), but are final for non-matches. For example,

- A match in IAFIS (of a ten-print fingerprint card) is final if the matcher score is especially high, but otherwise requires human verification. After the match is verified (if applicable), it is considered definitive proof of identity.
- A fingerprint match against a watchlist in US-VISIT starts a procedure of secondary processing, which may involve taking an additional set of fingerprints and running an IAFIS search.

- In the (now defunct) NCIC 2000 fingerprint system a fingerprint match was defined as probable cause for questioning; again, an IAFIS search would be conducted to determine identity.
- When the State Department checks visa applications using face recognition, any candidate matches require human verification.

Positive ID systems are final for matches (permitting access or use), but often have alternate processes for disputed non-matches, to enable access for falsely rejected legitimate users, or users with temporary limitations such as bandages. For example,

- Computers with biometric access control systems often require an alternate logon procedure (such as use of a smartcard or password).
- Physical access control systems often require security procedures to handle exceptions.

5.6 Human Verification

Human verification of results is a special type of end-stage matching process, but with some specific characteristics that differ from standard matchers:

- Human verification is slower than any automated process.
- Humans can make mistakes by accident in addition to errors of expertise/accuracy.
- Human identification is non-deterministic: an individual may give inconsistent responses for the same data.
- Humans are susceptible to bias, especially in terms of prejudging conclusions.
- The abilities of different people vary dramatically in terms of expertise, consistency, and susceptibility to mistakes and bias.

Engineering an intelligent fusion of matcher scores and human verification determinations is not straightforward. While automated matcher scores can in many or most cases be expressed in a relatively smooth distribution, human verification results for a single examiner are probably limited to a small number of discrete categories.⁷ At the very least, non-concurring results should be flagged, so that a high matcher score followed by a human non-match determination should be reviewed by another examiner.

Secondary processes, human verification, and exception handling are very costly in terms of throughput. In large-scale systems all secondary processes require consideration of load balancing, queuing issues, and tradeoffs between accuracy and throughput.

6 References

- [DataQuality] A. Hicklin and R. Khanna; "The Role of Data Quality in Biometric Systems"; February 2006.
- [FpVTE] C. Wilson, A. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Micheals, S. Otto and C. Watson; "Fingerprint Vendor Technology Evaluation 2003"; NIST Interagency Report 7123. June 2004.

⁷ At a minimum, {match | non-match | indeterminate}. Multiple levels of match or non-match certainty can be used, but a greater number of levels limits inter- and intra-examiner consistency.

- [Jain-00] A. Jain, R. Duin, J. Mao; "Statistical Pattern Recognition: A Review"; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, 2000, 4-37.
- [Jain-02b] Anil Jain and Arun Ross; "Fingerprint Mosaicking"; *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, Florida, May 13 - 17, 2002.
- [Korves-05] H. Korves, L. Nadel, B. Ulery, D. Masi; "Multi-Biometric Fusion: From Research to Operations"; *Sigma*, Summer 2005
- [Maltoni-03] Davide Maltoni, Dario Maio, Anil K. Jain, Salil Prabhakar; "Multimodal Biometric Systems", Chapter 7 in *Handbook of Fingerprint Recognition*; Springer; 2003; ISBN: 0387954317.
- [Ross-05] Arun Ross and Rohin Govindarajan; "Feature Level Fusion Using Hand and Face Biometrics"; *Proc. of SPIE Conference on Biometric Technology for Human Identification II*, Vol. 5779, pp. 196-204, March 2005.
- [SC37-24722] SC37 Working Draft Technical Report 24722 on "Multimodal and Other Multibiometric Fusion"; 2006-02-14 [ISO/IEC JTC 1/SC 37 N1506]
- [Senese-06] Francis Senese; "Livescan Records: A Potential Issue For The Fingerprint Community"; *NIST Latent Testing Workshop*, April 5 2006.
- [SlapSeg] B. Ulery, A. Hicklin, C. Watson, K. Kwong; "Slap Segmentation Evaluation 2004"; NIST Interagency Report 7209. March 2005.
- [StudiesOfFusion] B. Ulery, A. Hicklin, C. Watson, W. Fellner, P. Hallinan; "Studies in Biometric Fusion"; NIST Interagency Report 7346; September 2006.
- [Uludaga-04] Umut Uludaga, Arun Ross, Anil Jain; "Biometric template selection and update: a case study in fingerprints"; *Pattern Recognition* 37 (2004) 1533– 1542.
- [Wein-05] Wein and Bavega; "Using Fingerprint Image Quality to Improve the Identification Performance of the US-VISIT Program"; *Proceedings of the National Academy of Sciences*; May 24, 2005, vol. 102, no. 21.