

A Brief Introduction to Biometric Fusion

16 June 2006

Austin Hicklin,¹ Brad Ulery,¹ Craig Watson²

¹Mitretek Systems

²National Institute of Standards and Technology

Abstract

Biometric fusion is the use of multiple biometric inputs or methods of processing to improve performance. The key purposes for biometric fusion are to improve system accuracy, efficiency, applicability, and robustness. Some types of fusion have been used successfully for years in large-scale fingerprint identification systems. While fusion can be very effective, it should not be regarded as a panacea, since it adds complexity to data collection and system architecture.

Contents

1	What is biometric fusion?.....	3
1.1	Multi-Biometric Systems	3
1.1.1	Categories of fusion	3
1.1.2	Levels of fusion.....	4
1.2	Purposes for fusion	4
1.3	Limitations of biometric fusion	5
2	How do decision- and score-level fusion work?.....	5
2.1	Matcher scores and decisions	5
2.2	Matcher score distributions	6
2.3	Decision- and score-level fusion	7
2.4	Can a less accurate matcher improve on a more accurate matcher?.....	8
2.5	Criticisms of fusion	9
3	Conclusion.....	10
4	References.....	10

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 a number of colleagues for their reviews and comments on earlier drafts of this document, including Elham Tabassi (NIST); George Kiebusinski, Larry Nadel, and Shahram Orandi (Mitretek); and Harris Ulery.

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 What is biometric fusion?

Biometric systems are automated means by which physical traits (or sometimes behavior) are used to identify a person, or verify a person's identity. A variety of such systems have been implemented and used successfully over the years, including ones based on fingerprints, irises, facial images, hand geometry, and speaker recognition, among others. The successful implementation of biometric systems requires addressing a number of issues, including accuracy, efficiency, robustness, applicability, and universality.

One method of dealing with many of the issues confronting biometric systems is to collect more data from each subject, and fuse the data, or the results of processing that data. Biometric fusion can be defined broadly as the use of multiple types of biometric data or methods of processing to improve the performance of biometric systems.

The theory behind fusion is not limited to biometrics: biometric-based decisions are a special case of classification in the field of statistical pattern recognition, and biometric fusion analogously can be considered a special case of combining multiple classifiers in pattern recognition. Fusion methods are used in such diverse fields as Internet search engines, analysis of satellite imagery, and analysis of medical test results.

Biometric fusion is not a new idea: for years, various aspects of fusion have been integral part of the successful implementation of biometric systems, especially large-scale fingerprint systems.

1.1 Multi-Biometric Systems

A *multi-biometric system* is one in which multiple categories of data are collected and used for various purposes, including but not limited to fusion:

- *Selection*, in which the best or most useful data is retained for use, while the other data is ignored or discarded. Selection is frequently based on quality metrics.
- *Validation*, in which some of the data is used to check the integrity of the other data.
- *Fusion*, which is based on combining data at various levels.

Biometric fusion is generally classified in terms of both *categories* and *levels* [SC37-24722]. The categories define what inputs or processes are being used for fusion; the levels define how the fusion performed.

1.1.1 Categories of fusion

The types of data or methods of processing used constitute the *categories* of fusion:

- *Multi-sample*: fusion of multiple samples (images) acquired from the same source, such as multiple images of a single fingerprint, images of the same face, or recordings of a speaker.
- *Multi-instance*: fusion of multiple instances of the same type of biometric, such as fingerprints from multiple fingers, or images of both irises.
- *Multi-modal*: fusion of multiple types (or modalities) of biometrics, such as a combination of a subject's fingerprints, face, irises, and voice.
- *Metadata*: fusion of biometric inputs with other information, such as measures of sample quality, or demographic information such as gender, height, or age. Demographic information is sometimes described as soft biometrics.

- **Multi-algorithm:** fusion of multiple methods of processing for each individual sample. In practice, this usually means the use of multiple matchers, but can also apply to multiple methods of feature extraction.

1.1.2 Levels of fusion

As shown in Figure 1, there are several stages in the process of arriving at an identity decision. A sample (image) is converted in feature extractor software into a template (machine representation, feature set). Matchers compare the template against templates from the gallery (database), and generates a system-specific matcher score (similarity score). Decision software compares the matcher score to predetermined thresholds to make a match or non-match determination.

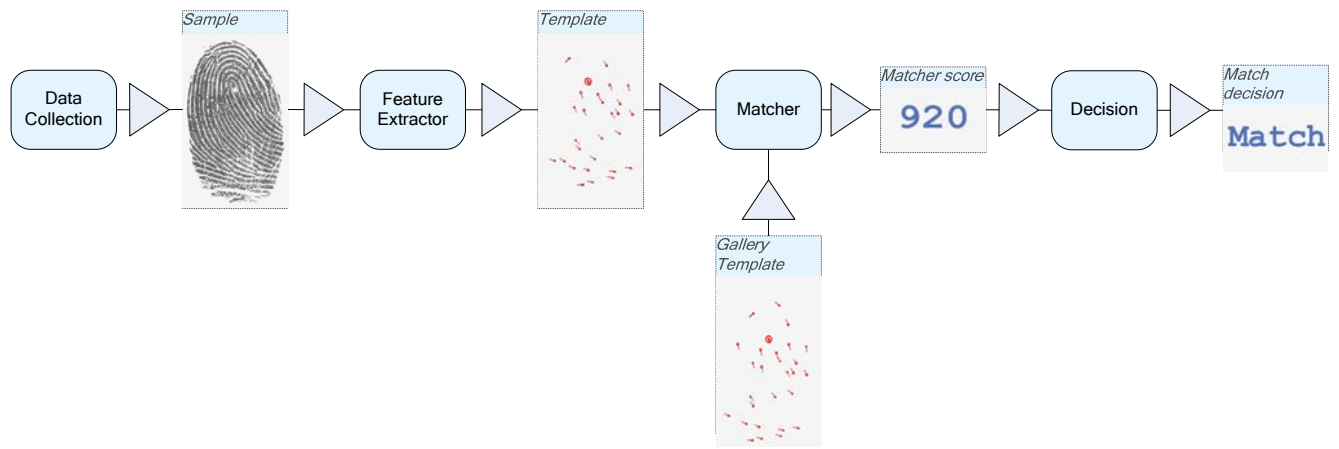


Figure 1: Stages of processing in a simplified biometric system.

The means by which data can be fused are known as the *levels* of fusion, which correspond to the stages of processing:

- **Sample-level fusion:** For multi-sample systems, the samples can be combined to form a single sample. For example, livescan fingerprint devices create a single large “rolled” image from a series of individual fingerprints.
- **Template-level fusion:** Feature extractor software converts samples (images) into simplified computer representations known as templates or feature sets. In template-level fusion, multiple templates are combined to form a single template.
- **Score-level fusion** refers to methods in which multiple samples, instances, or modalities are compared, and the resulting similarity scores (or probabilities) are combined to form a single fused score. Score-level fusion can also be used to combine the results of multiple algorithms when a single sample is searched.
- **Decision-level fusion** is used in the same cases as score-level fusion, but the scores are turned into match/non-match decisions before fusion.

1.2 Purposes for fusion

Fusion has been used successfully for years in large-scale automated fingerprint identification systems (AFIS), which combine multi-finger data and multiple methods of processing; indeed, it is fair to say that multi-instance and multi-algorithm fusion are what have made very large-scale

fingerprint systems practical. Today, various forms of fusion are used in a number of different types of biometric systems.

Fusion can be used to address a number of issues faced by the designers, implementers, and operators of biometric systems:

- **Accuracy:** Fusion can be used very effectively to improve overall accuracy. Biometric system accuracy is generally stated in terms of maximizing the True Accept Rate¹ (TAR) while minimizing the False Accept Rate (FAR): maximizing the ability to recognize those subjects who have already been enrolled, without incorrectly identifying them as other subjects.
- **Efficiency:** Fusion can be used to increase efficiency, or to allow tradeoffs between efficiency and accuracy. System efficiency can be stated in terms of throughput (processing time), computational requirements, and financial cost.
- **Robustness:** The inherent redundancy in a fused system increases the system's robustness. Robustness refers to the ability of a system to continue to function as accurately as possible despite problems such as poor sample (image) quality and data integrity errors.
- **Applicability:** Applicability relates to the appropriateness of a system for a task: the need to work with legacy data often dictates the biometric modalities that can be used. A multi-modal system is more applicable to a broad variety of uses than a uni-modal system, because it can be used in conjunction with multiple sources of legacy data. For example, a multi-modal fingerprint and face system can conduct both fingerprint-only background checks and face-only watchlist checks.
- **Universality:** Universality refers to whether all people can use a given biometric system. Some people cannot provide usable biometric samples, for reasons such as amputations, injury, or disease. Multi-modal and multi-instance systems can provide alternatives so that all people can use a system.

1.3 Limitations of biometric fusion

Of course, fusion comes at a price. Collecting additional data takes time, adds complexity and cost to the collection process. Collection of an increased amount of biometric data is likely to increase public concerns about privacy issues and intrusiveness. The software and hardware required for additional processing adds both complexity and cost to the system.

2 How do decision- and score-level fusion work?

2.1 Matcher scores and decisions

Every comparison of two samples processed in a matcher results in some measure of similarity or difference, which is known as a matcher score. Biometric systems make match or non-match decisions based on whether those matcher scores exceed an operationally specified threshold.

¹ See the Glossary for definitions of biometric terminology.

2.2 Matcher score distributions

The distributions of matcher scores for known genuines and imposters² can be graphed, as shown in Figure 2. In a simple system, these matcher scores are compared to a threshold value, and comparisons are determined to be matches if they are above that threshold. In Figure 2, a comparison with a score above 0.65 would be very likely to be a match (because little of the red imposter distribution is to the right of that point), but that decision criterion would exclude many true genuines (because much of the black genuine distribution is to the left of that point). In other words, given a threshold of 0.65, the false accept rate would be low, but the false reject rate would be very high. A lower threshold would decrease the number of falsely rejected genuines, but at the cost of an increase in the number of falsely accepted imposters.

Operational thresholds are set according to the needs of each application: most large-scale identification systems must have extremely low false accept rates, at the cost of increased false rejects; many physical access control systems must have low false reject rates, at the cost of increased false accepts. An equal error rate (EER) is the score threshold at which the false accept and false reject rates are equal.

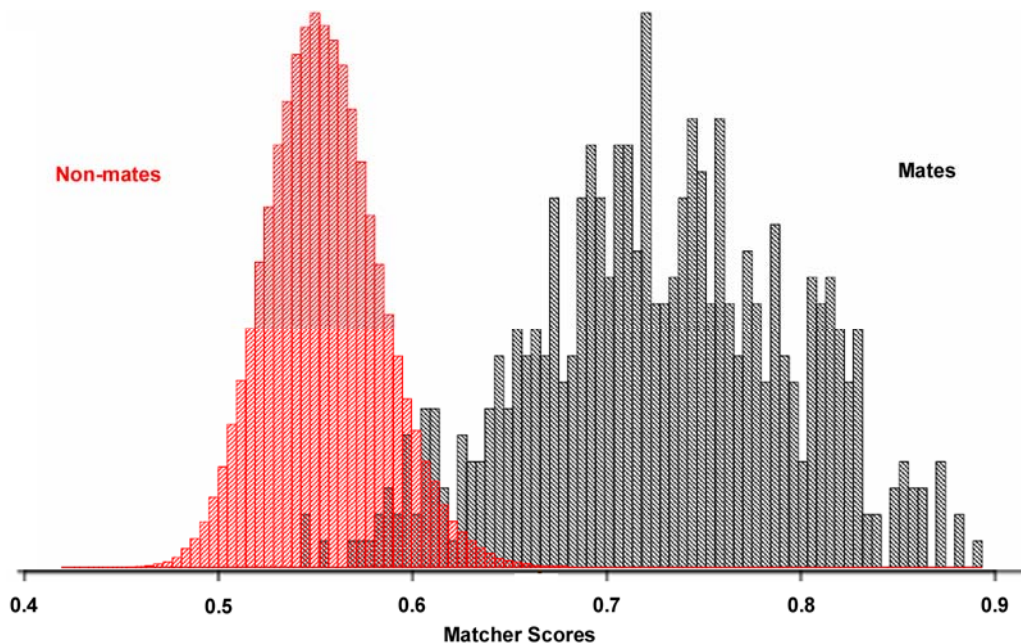


Figure 2: Example of **imposter (red)** and **genuine (black)** matcher score distributions for a moderately accurate matcher. Note that the scores for imposters are lower than the from scores of genuines, but the distributions overlap. More accurate matchers show a higher degree of separation. Most matcher scores are matcher-specific, and so must be normalized before being compared between matchers. *[BSSR1 dataset; face G matcher]*

² A system makes a *match* or *non-match* decision based on a score threshold; *genuine* or *imposter* refer to whether the samples actually came from the same individual. Evaluations are based on measuring the differences between the “ground truth” (genuines and imposters) and the system decisions (match and non-match). In operational systems, knowing the ground truth may not be possible.

2.3 Decision- and score-level fusion

When independent matcher score distributions are plotted against each other in an X-Y scatterplot, the resulting graph shows a much greater separation between imposter and genuine distributions than the graph of either distribution alone. This can be seen in Figure 3; note that the matcher in Figure 2 is shown in the Y axis.

This figure shows the distributions of face and fingerprint matcher scores from a sample dataset (the NIST BSSR1 data; datasets are described in *IV: Description of Datasets and Pre-Fusion Data Characteristics*). Right index fingerprint scores from the NIST-VTB matcher are on the X axis and face scores from the anonymous C matcher are on the Y axis. Note that neither of these matchers is state of the art today.

The gray shading of Figure 3 represents decision-level fusion: the set of all shaded areas (light or dark gray) represent those accepted by an OR decision (above the X threshold OR above the Y threshold); those in the dark gray region satisfy an AND decision criterion. The thresholds for the gray regions are set at a low false accept rate, such as might be appropriate for an identification system. Note how the gray regions provide much better discrimination of genuines and imposters than is possible with either matcher alone. For comparison, the blue lines show thresholds based on equal error rates.

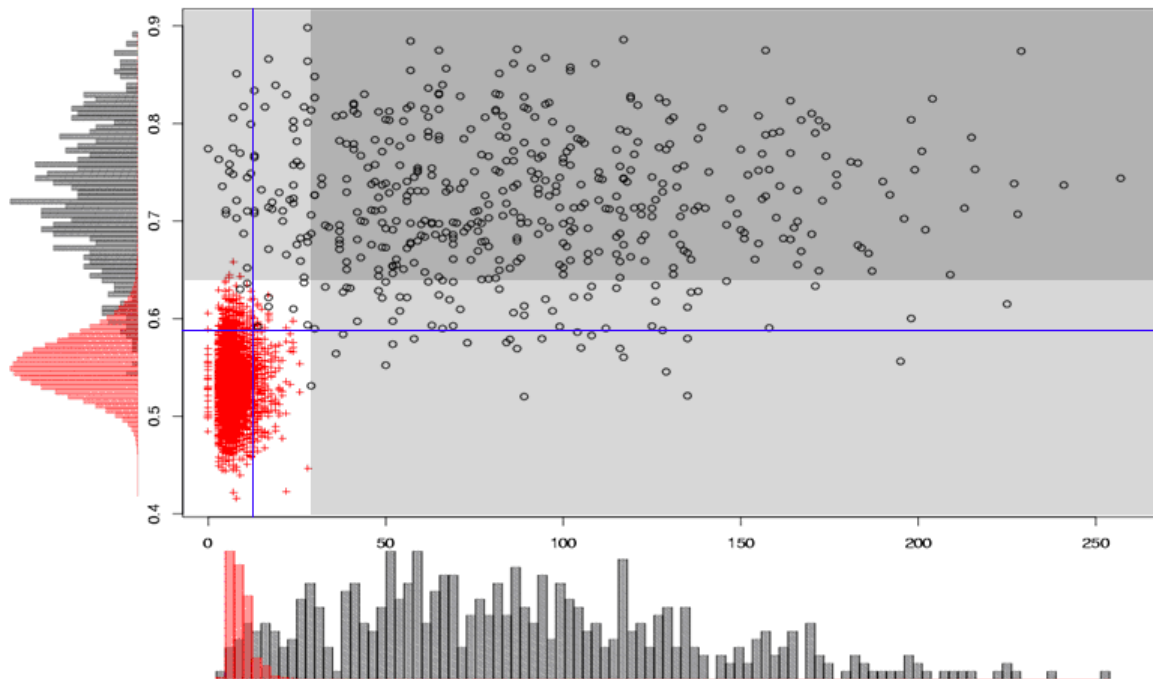


Figure 3: Example of combined imposter (red) and genuine (black) matcher score distributions for two matchers. The gray areas show decision-level fusion thresholds. The dark gray area is AND rule decision fusion; the light and dark gray areas together represent OR rule decision fusion. The blue lines show thresholds based on equal-error rates.

[BSSR1 dataset; fingerprint V & face G matchers]

Figure 4 shows the same score data as Figure 3, but shows how the decision boundary can be improved with access to matcher scores rather than relying exclusively on matcher decisions. This example shows linear fusion boundaries as well a polynomial boundary, each of which is

superior to the decision-level methods in Figure 3. More elaborate — and accurate — decision boundaries follow the topology of the intersecting genuine and imposter distributions.

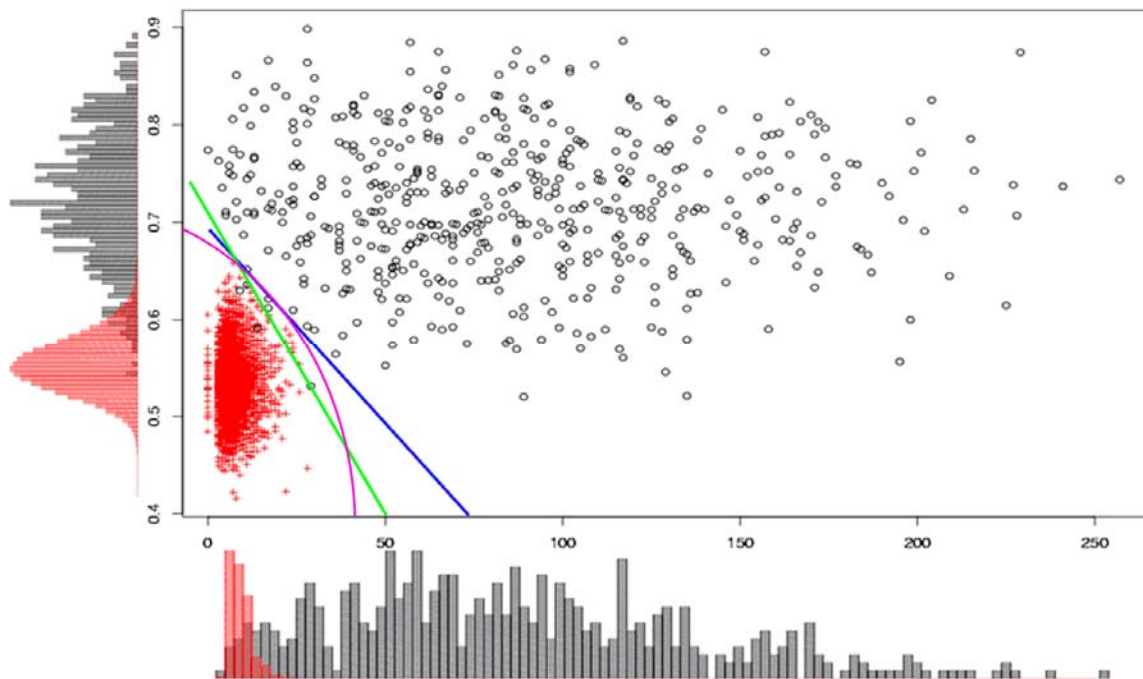


Figure 4: Example of combined **imposter** (red) and **genuine** (black) matcher score distributions for two matchers. The lines and curve are examples of score-level fusion decision boundaries.

[BSSR1 dataset; fingerprint *V* & face *G* matchers]

2.4 Can a less accurate matcher improve on a more accurate matcher?

Even in the case of a highly accurate matcher, fusion with additional data and/or other matchers *will* improve accuracy if that addition contributes useful information, and fusion is correctly implemented. Figure 5 shows an example of how a less-accurate face matcher can be used to improve on a more-accurate single-finger matcher. The histogram at the bottom shows that most of the fingerprint genuine scores are extremely high (over 80% of scores are at the maximum score value, at the right side of the chart), but a relatively small proportion of genuines are spread across the range of scores, some of which overlap the imposter scores. While the overall distributions of the face genuine and imposter scores are not as distinctly separated, they can be valuable when used in combination with the fingerprint scores.

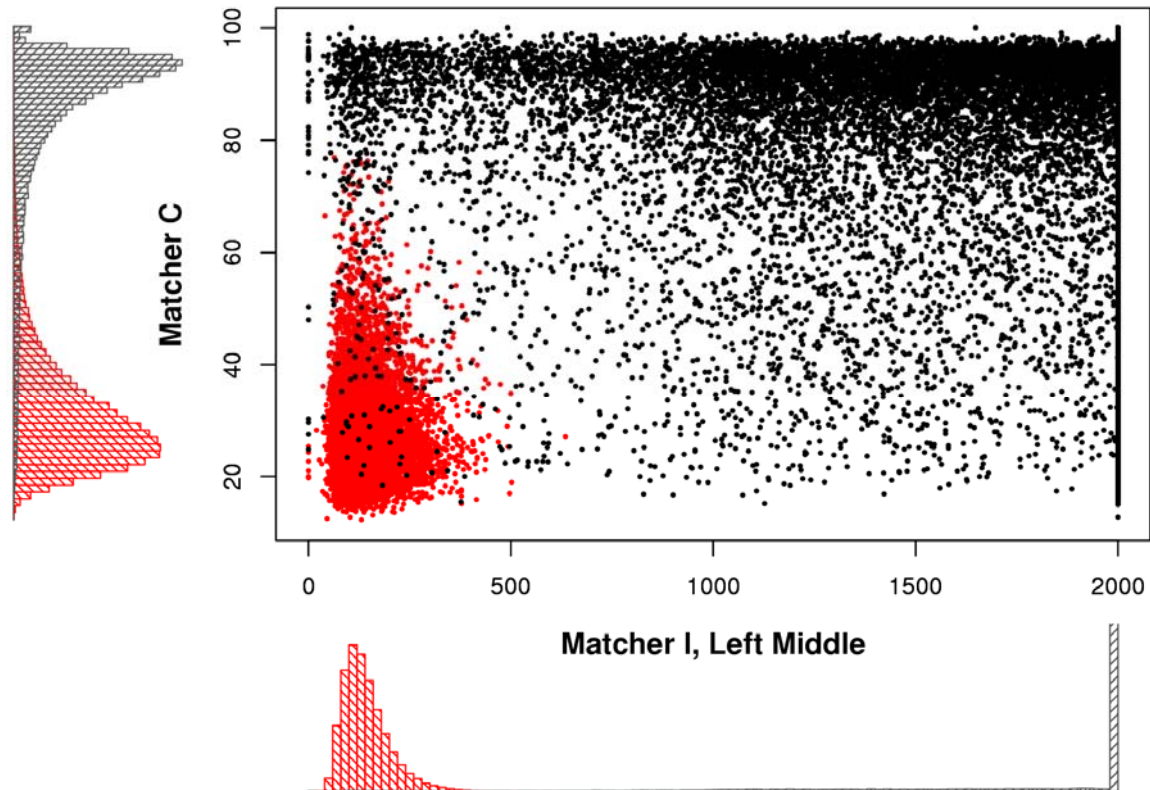


Figure 5: Example of how a less accurate face matcher (Y axis) can improve system performance when fused with the more accurate fingerprint matcher (X axis). Note that most of the genuines (black) with low fingerprint scores can be separated from the imposter (red) distribution by use of face scores. (approx. 65,000 genuines and 122,000 imposters; known data integrity errors are excluded from this chart)

[NBDF06 dataset, 33 data integrity errors excluded]

2.5 Criticisms of fusion

Some recent articles have criticized biometric fusion, such as a *Register* article entitled “Read two biometrics, get worse results” [Lettice], or a London School of Economics report, “The Identity Project: Research Status Report”, which stated “Combining biometrics is both theoretically and practically challenging with dubious results.” [LSE-Status, p.28]

Both of these cite a short paper by John Daugman, “Combining Multiple Biometrics.” [Daugman] This paper shows the weaknesses of decision-level fusion of biometrics, but his analysis is based on a particularly ineffective method: selecting decision thresholds that are both set at equal-error rates (EER). Decision-level fusion is known to be suboptimal, but can be useful if thresholds are set properly. The paper proceeds to show that other thresholds improve performance, and states that the results only apply to decision-level fusion, but does not note these critical elements in the overview or the bold conclusion. Unfortunately, these subtleties were completely ignored in the London School of Economics report, and while the caveats were mentioned in the text of the *Register* article, they did not prevent a misleading headline.

3 Conclusion

Biometric fusion is defined broadly as the use of multiple types of biometric data or methods of processing to improve the performance of biometric systems. Fusion works by combining information from multiple sources. This is done to improve the accuracy, efficiency, and robustness of biometric systems. We have shown specific examples of multi-modal, score- and decision-level fusion.

This paper is simply an introduction to biometric fusion, and purposely does not address the various ways that fusion can be used, the complexities of operational uses of fusion, technical explanations of the various methods of implementing fusion, and does not include extensive empirical results. For more detail in these areas, please see the companion papers in this collection.

4 References

- [Daugman] John-05] J. Daugman; "Combining Multiple Biometrics"; (No date)
<http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html>
- [Lettice-05] Lettice; "Read two biometrics, get worse results" The Register, Wednesday
19th October 2005. ;
http://www.theregister.co.uk/2005/10/19/daugman_multi_biometrics/
- [LSE-Status] London School of Economics and Political Science; "The Identity Project
Research Status Report"; January 2006. ;
<http://is2.lse.ac.uk/IDcard/statusreport.pdf>
- [SC37-24722] SC37 Working Draft Technical Report 24722 on "Multimodal and Other
Multibiometric Fusion"; 2006-02-14 [ISO/IEC JTC 1/SC 37 N1506]
[http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/2262372/2263034/
2300190/JTC001-SC37-N-1506.pdf?nodeid=4446319&vernum=0](http://isotc.iso.org/livelink/livelink/fetch/2000/2122/327993/2262372/2263034/2300190/JTC001-SC37-N-1506.pdf?nodeid=4446319&vernum=0)