

When to Fuse Two Biometrics

Elham Tabassi, George W. Quinn, and Patrick Grother
National Institute of Standards and Technology
Gaithersburg, MD USA

{tabassi, gwquinn, pgrother}@nist.gov

Abstract

Biometric fusion, acquisition and combination of multiple pieces of evidence of identity, can achieve higher accuracy of biometric recognition than using a single biometric. However, fusion increases the cost or throughput of the system since it requires acquisition and processing of more samples. We document a procedure for contingent fusion. That is, two biometrics are fused only if verification on the first presented biometric is rejected. We present results of this approach for both decision and score level fusion, and examine the combination of two different algorithms, two different modalities, and two different instances of a biometric. We conclude that contingent fusion results in comparable accuracy and lower cost as measured by processing time per sample than always fusing two pieces of evidence of the same identity.

1. Introduction

Several studies [1, 2, 3] have shown that consolidating information from multiple biometric sources can significantly enhance the accuracy of a biometric system. Further, employing multiple sources can alleviate problems with noisy data, high failure to enroll rates, and can make it more difficult for an intruder to violate the systems. Multiple biometric sources can be any combination of multiple sensors (e.g. optical or solid state), multiple biometric traits (e.g. finger, iris), multiple representations (e.g. 3D and infrared face), multiple instances (e.g. left and right index fingers), multiple samples of the same biometric (e.g. frames of a video sequence of face), or multiple matching algorithms [4]. Information from these sources can be integrated at the feature, score, or decision level.

Biometric fusion improves the accuracy of a biometric recognition system, and it is necessary in cases like missing biometrics (like amputated fingers), but often leads to additional cost in complexity, processing time, or resources. For example, fusing multiple traits of biometric or multiple representation of a biometric, usually requires deploy-

ment of different sensors, which obviously adds to the complexity of the system. Having to capture two different biometrics (e.g. face and finger) increases the capture time per user. Besides, recent evaluations of fingerprint recognition algorithms [5] have shown that the most accurate fingerprint recognition systems can correctly verify a claim of identity with a false non-match rate of 0.01 or better at a false match rate of 0.001. In these cases, fusion could increase the cost without tangible improvement in accuracy. Although there are many publications on *how* to fuse biometrics [1, 2, 3], to our knowledge, there are no studies addressing the question of *when* to fuse biometrics. In this paper, we address the latter question by proposing contingent fusion. We focus on the concept of conditionally combining information from another biometric source or matching algorithm to maximize improvement in accuracy while minimizing the increase in cost. We study cost versus performance for multi-instance, multi-algorithm, and multi-modal contingent fusion and compare them with both singular no-fusion case and the always-fuse case. For each of these cases, we perform quality-based fusion as well. We quantify cost as transaction and processing time per sample and performance as false match rate of the fused system.

The rest of the paper is organized as follows: we propose the concept of contingent fusion in section 2, present experimental results in section 3, and follow with some conclusions in section 4.

2. Contingent fusion

The contingent fusion scenario we proposed is a two-stage process: two independent biometric tests are combined by disjunctive (*OR*) rule. An identity claim is accepted if the first test is passed so only those who were rejected in the first test are subject to the second test. That is, a false non-match can only happen if both biometric tests result in a false non-match. Furthermore, false match is the complement of the probability that neither test₁ nor test₂ results in a false match [6].

To calculate the total false non-match rate (FNMR_T) and false match rate (FMR_T) of the fused system, let FNMR_i

and FMR_i denote false non-match rate and false match rate of test $_i$ respectively where $i = 1, 2$. Therefore

$$FNMR_T(\tau_1, \tau_2) = FNMR_1(\tau_1) \cdot FNMR_2(\tau_2) \quad (1)$$

$$FMR_T(\tau_1, \tau_2) = 1 - (1 - FMR_1(\tau_1)) \cdot (1 - FMR_2(\tau_2)) \quad (2)$$

It is obvious that $FNMR_T$ is lower and FMR_T is higher than for either test alone.

We define cost as the average processing time per user as did Wein *et al.* in [7]. Although legitimate users can be correctly accepted at either the first or second test, those rejected at the first test will require additional processing time to perform the second test. Legitimate users falsely rejected by both tests will also require additional processing time for secondary (often by human) inspection. Therefore, the overall average processing time per legitimate user can be modeled as:

$$C(\tau_1, \tau_2) = c_1 + c_2 \cdot FNMR_1(\tau_1) + c_3 \cdot FNMR_T(\tau_1, \tau_2) \quad (3)$$

where c_1 , c_2 , and c_3 are transaction and processing time per sample for test $_1$, test $_2$, and secondary inspection respectively, and τ_i is threshold value for test $_i$, $i = 1, 2$, but the procedure can be generalized to $i \geq 3$. By defining the cost function in terms of time, instead of the $FNMR_T$ only, we are able to factor in the time requirements involved with acquiring and processing one or both biometrics per user. Ideally, one would like to have both time per user and FMR_T as low as possible, but realistically one has to find a balance between the two. Since the FMR_T is bounded by FMR_1 , the threshold for test $_1$ (τ_1) should be set high enough to ensure acceptable FMR_T . However, forcing τ_1 to be large results in higher $FNMR_1$ which means an increase in cost $C(\tau_1, \tau_2)$. Plots of Figure 1 show the trade-offs of $FNMR_T$ and FMR_T vs. the two thresholds of test $_1$ and test $_2$ where test $_1$ is verifying the right index finger and if rejected, left index finger of the same person is presented to the same matching algorithm (test $_2$). Scores of the right and left index fingerprints are added together and the claim of identity is accepted if the sum of scores are bigger than the operating threshold and rejected otherwise. (This is multi-instance contingent sum fusion as explained in section 2.1).

Transaction and processing time per image consists of non-biometric (average time users take to position themselves by the sensor plus the acquisition time) and biometric time (average time matching algorithms take to process, compute quality score if applicable, and match an image). That is,

$$c = c_{non-biometric} + c_{biometric} \quad (4)$$

where

$$c_{non-biometric} = t_{placement} + t_{acquisition} \quad (5)$$

and

$$c_{biometric} = t_{extraction} + t_{matching} + t_{quality} \quad (6)$$

Both biometric and non-biometric costs are different for different modalities, different algorithms, and might even differ based on the quality of the biometric sample, for example an intelligent matching algorithm might invoke some image enhancement procedures for poor quality images. In this paper we assume that biometric processing time is the same for all images of the same modality regardless of their quality.

Usually the non-biometric time (equation 5) is much larger than the biometric time (equation 6). It takes about five seconds per finger to capture a digital image of a fingerprint [7]. We estimate average time users need to position themselves by a sensor as two seconds, and hence non-biometric cost per user (*i.e.* $c_{non-biometric}$ in equation 5) as 7 seconds. According to [5], template generation and match time for fingerprint images are different for different vendors, but mostly about half a second. Computing quality values for fingerprint images takes about 0.3 (for flat impressions of fingerprints) to 0.5 (for rolled impressions of fingerprints) second [8]. Therefore, biometric cost $c_{biometric}$ in equation 6 is estimated as 0.5 second for a typical matcher when quality is not computed and 1 second when quality of the image is computed by NFIQ for quality-based fusion. That means we estimated the average total non-biometric and biometric cost of a typical matcher per fingerprint as 7.5 seconds if quality is not computed, and 8 seconds for quality-based fusion scenarios.

For face recognition systems, we assume a total of 20 seconds non-biometric and biometric time as it may take more time for users to position themselves in front of a camera than a fingerprint scanner, plus capture time and biometric processing time might be longer than fingerprint.

We chose c_3 , the time a human needs to check the claim of identity for the user rejected by both tests, to be 15 minutes. Although we chose values for c_1 , c_2 , c_3 deliberately to be representative of real operational systems, nevertheless, these numbers are just estimates. In an operational scenario, c_1 , c_2 , c_3 should be tailored to the intended application, as these values will have a big effect on the cost vs. FMR_T tradeoff conclusions.

Our performance measure involves comparing the trade-off between the FMR_T and the transaction and processing time per user. Each scenario in our model involves solving an optimization problem in which we attempt to

$$\begin{aligned} & \text{minimize } C(\tau_1, \tau_2) \\ & \text{subject to } FMR_T \leq f \end{aligned} \quad (7)$$

where $C(\tau_1, \tau_2)$ and FMR_T are defined in equations 3 and 2 respectively, and f is the maximum allowable false match

rate. This is achieved by surveying the two thresholds τ_1 and τ_2 (used at the first and second tests) to get the lowest cost while still maintaining an acceptable overall false match rate. (False match rate is a measure of security and what constitutes acceptable false match rate is application dependent.)

We consider three cases for the two biometric tests:

1. *Multi-instance fusion* - two different instances of the same biometric (*i.e.* right and left index fingerprints),
2. *Multi-algorithm fusion* - two different matching algorithms of the same biometric (*i.e.* fingerprint),
3. *Multi-modal fusion* - two different biometric traits (*i.e.* face and fingerprint).

We combine scores of test_1 and test_2 , for each above mentioned case, according to the following fusion schemes:

1. *Decision level fusion*: If test_1 rejects the first sample (*i.e.* flat impressions of right index), the other sample (*i.e.* flat impressions of left index or face) is presented to test_2 which either accepts or rejects the claim of identity solely based on the second sample. (For the multi-algorithm case the same impression of right index finger is presented to both tests.)
2. *Sum score fusion*: If test_1 rejects the first sample (*i.e.* right index), the other sample (*i.e.* left index or face) is presented to test_2 . (Again for the multi-algorithm case the same impression of right index finger is presented to both tests.) Final decision (accept or reject the identity claim) is based on the sum of scores

$$s_{fused} = s_1 + s_2 \quad (8)$$

where s_i is score produced by test_i and $i = 1, 2$. Kittler *et al.* [1] have shown that sum fusion, though very simple, is a reasonable way of combining genuine scores of multiple tests when appropriate score normalization is applied [10].

3. *Quality based log likelihood fusion*: We used NIST Fingerprint Image Quality (NFIQ) [8, 9] to compute quality values for fingerprint images. Then we empirically computed genuine and impostor distribution for each of the five levels of NFIQ, which are shown in Figure 2. Scores of test_1 and test_2 are combined by likelihood ratio, where the fused score is

$$S_{fused} = \log \frac{m_1(s : q = j)}{n_1(s : q = j)} + \log \frac{m_2(s : q = k)}{n_2(s : q = k)} \quad (9)$$

$m_i(s : q = j)$ and $n_i(s : q = j)$ are genuine and impostor probability density functions of test_i , ($i = 1, 2$) of samples with quality j , empirically computed,

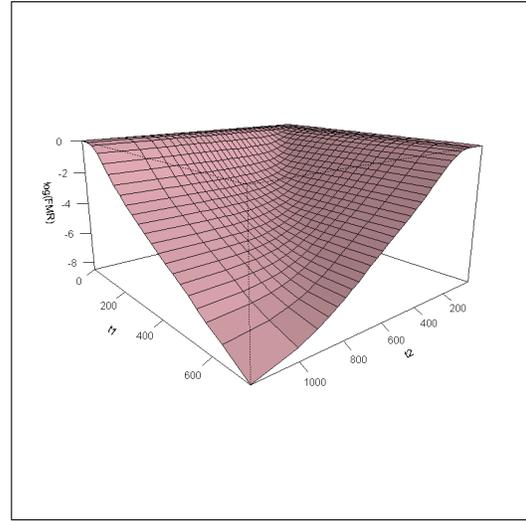
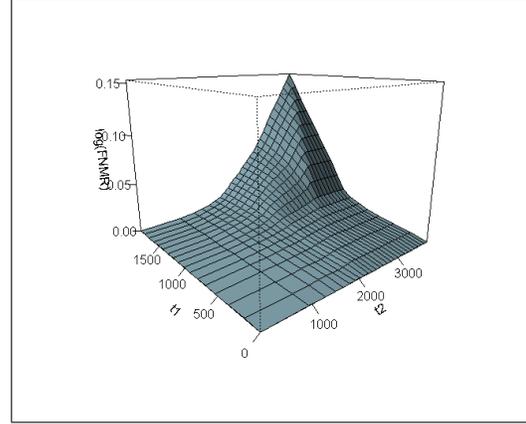


Figure 1. Surface plots of (a)FNMR and (b) $\log(FMR)$ of the multi-instance decision level fused system vs. thresholds of test_1 and test_2 . The higher the thresholds the lower the FMR_T , but the higher $FNMR_T$.

while j , and $k = 1, \dots, 5$ correspond to the qualities of the first sample (presented to test_1) and second sample (presented to test_2) respectively. Note that $q = 1$ is the highest quality level and $q = 5$ is the lowest.

In each case, we iterate over all possible threshold values of test_1 and test_2 and choose the minimum cost (equation 3) for each $FMR_T \geq 0.0001$, which are plotted for the above mentioned cases. We used commercial fingerprint and face matching algorithms and data collected at operational environments.

We now discuss each of these cases in more detail.

2.1. Multi-instance contingent fusion

In this scenario, we fuse multiple instances of a biometric *i.e.* left and right index fingerprints. The right index fin-

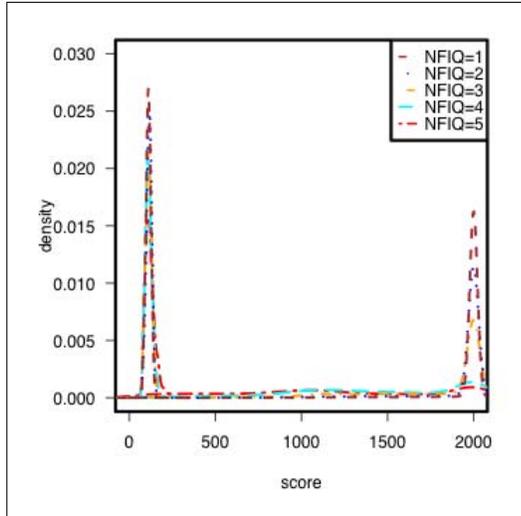


Figure 2. Genuine and impostor distributions of NFIQ five quality levels.

gerprint of a user is presented to test₁ and if rejected, the left index fingerprint image is presented to test₂. The most common use-case of this scenario is an access control scenario like [12].

Since the same matching algorithm is being used for test₁ and test₂, c_1 and c_2 of equation 3 are equal. Therefore, the cost of multi-instance contingent fusion can be written as:

$$C_C(\tau_1, \tau_2) = c_1 + c_1 \cdot \text{FNMR}_1(\tau_1) + c_3 \cdot \text{FNMR}_T(\tau_1, \tau_2) \quad (10)$$

which is much smaller than the cost for always fusing two instances,

$$C_A(\tau_1, \tau_2) = 2c_1 + c_3 \cdot \text{FNMR}_T(\tau_1, \tau_2) \quad (11)$$

where $c_1 = 8$ seconds for quality-based log likelihood ratio fusion scheme and 7.5 seconds for all the other fusion schemes, as explained in section 2, C_C is the cost for contingent fusion and C_A is the cost for always fusing two instances.

Figure 3 shows cost vs. FMR_T for the single-instance case (when only one fingerprint is authenticated), always performing both tests and sum fusing the scores, and three contingent fusion schemes explained in section 2. For FMR_T of 0.001, contingent fusion reduces the cost by 50% for single-instance and 30% for always-sum fusing right and left index fingers. Furthermore, as shown in Table 1, contingent fusion systems' FNMR_T at $\text{FMR}_T = 0.001$ is an order of magnitude smaller than single-finger case. Performance is quite comparable for quality based contingent log likelihood ratio and contingent sum fusion schemes which agrees with Jain *et al.* [11] findings.

Table 1. Multi-instance contingent fusion. FNMR_T and cost are computed at $\text{FMR}_T = 0.001$.

fusion scheme	FNMR_T	cost (seconds)
single-instance	0.0136	19.7
contingent decision	0.0019	9.3
contingent sum	0.0016	9.1
contingent L-ratio	0.0017	9.7
always sum	0.0016	16.5

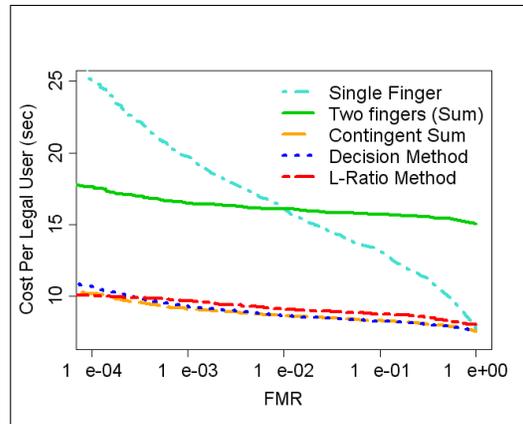


Figure 3. Multi-instance contingent fusion

2.2. Multi-algorithmic contingent fusion

This scenario involves fusing results of two algorithms applied to the same biometric sample (*i.e.* right index) presented to both algorithms. It is often the case that an accurate matching algorithm takes more time to process and compare two samples than a less accurate algorithm. The cost associated with capture (non-biometric cost) and process (biometric cost) for test₁ is 7.5 second as explained in section 2. However, there is no non-biometric cost for test₂ (because there is no transaction involved), and since we assumed test₂ deploys a more accurate but slower algorithm than test₁, biometric cost of test₂ is estimated at 1.5 second (compared with half a second as mentioned in section 2), *i.e.* $c_2 = 1.5$ second.

We consider the following fusion scenarios:

1. *Decision level fusion* - We model decision level contingent fusion by first presenting the right index sample of a user to the matching algorithm₁ (test₁) and if rejected, it is presented to the matching algorithm₂ (test₂), which is assumed to be more accurate but slower. Cost function for this scenario is same as equation 3 with $c_1 = 7.5$ (which includes non-biometric and biometric costs) and $c_2 = 1.5$ (which is only biometric cost of test₂) seconds.

2. *Quality-based decision fusion* - We also consider the case where fingerprint samples with NFIQ= 4,5 are sent directly to test₂. Such poor quality samples have a low likelihood of being recognized correctly [9] and so tend to fail test₁. Given our assumption that test₂ deploys a more accurate (and often) more intelligent algorithm, bypassing test₁ for such poor quality samples will reduce the cost. Samples with NFIQ=1,2,3 are presented to test₁ and if rejected to test₂, with cost function as in equation 3 with $c_1 = 8$ and $c_2 = 1.5$ seconds. Cost function for samples with NFIQ=4,5 is

$$C_C(\tau_2) = c_2 + c_3 \cdot \text{FNMR}_T(\tau_2) \quad \text{if NFIQ}=4,5 \quad (12)$$

where $c_2 = 9$ seconds, which is 7 seconds non-biometric cost, 0.5 second NFIQ computation, plus 1.5 extraction and match time by the slower but more accurate matcher.

3. *Contingent sum fusion* - We performed contingent sum as explained in section 2. Cost function for this scenario is same as equation 3 with $c_1 = 7.5$ (which includes non-biometric and biometric costs) and $c_2 = 1.5$ seconds (which is only biometric cost of test₂) for both cases of contingent and always sum fusing scores.
4. *Quality based log likelihood fusion* - If test₁ rejects identity claim, same sample is presented to test₂ and the scores of test₁ and test₂ are fused according to equation 9. Cost for this scenario is same as equation 3 with $c_1 = 8$ and $c_2 = 1.5$ seconds.

We compared these cases with single-algorithm case of using only one algorithm when either algorithm is used and always performing both tests and sum fuse scores where the cost is

$$C_A(\tau_1, \tau_2) = c_1 + c_2 + c_3 \cdot \text{FNMR}_T(\tau_1, \tau_2) \quad (13)$$

where $c_1 = 7.5$ and $c_2 = 1.5$ seconds (same as multi-algorithm contingent sum fusion).

Figure 4 shows cost *vs.* FMR_T for all the above mentioned cases, and Table 2 shows FNMR_T and cost at $\text{FMR}_T = 0.001$. Multi-algorithm fusion is not as effective as multi-instance fusion and requires deployment and maintenance of two algorithms which is more costly than deploying only one algorithm. Our results suggest that the use of the better matcher yields the best performance. An effective and successful multi-algorithm fusion depends on the appropriate choice of matchers and their operational thresholds.

2.3. Multi-modal contingent fusion

We also performed experiments when test₁ and test₂ are different biometric traits *i.e.* face and finger. Since commer-

Table 2. Multi-algorithm contingent fusion. FNMR_T and cost are computed @ $\text{FMR}_T = 0.001$.

fusion scheme	FNMR_T	cost(seconds)
single-algorithm - test ₁	0.0136	19.7
single-algorithm - test ₂	0.0082	15.8
contingent decision	0.0074	14.2
quality-based	0.0078	15.1
contingent sum	0.0093	15.9
contingent L-ratio	0.0068	14.1
always sum	0.0092	17.3

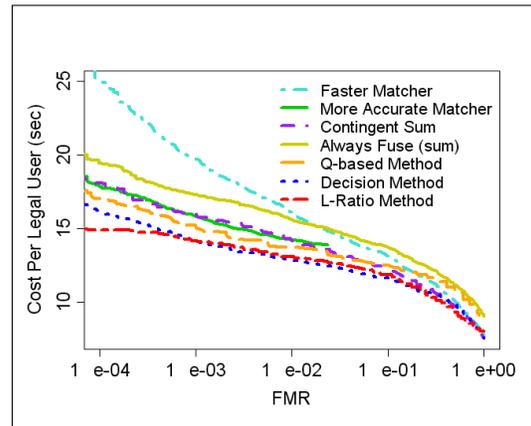


Figure 4. Multi-algorithm contingent fusion

cial fingerprint matchers offer more accuracy than commercial face recognition algorithms, we perform test₁ on fingerprint samples. Performing test₁ on fingerprint decreases the cost C by reducing FNMR_1 . If test₁ rejects the claim of identity presented to it, a face image is presented to test₂. Ideally we like to use face and fingerprint of the same person, but due to limitation of availability of data, we matched up fingerprint and face of different persons. Given that face and fingerprint of a person is assumed to be independent, matching face and fingerprint of different people should not have any adverse effect on the results. We performed decision fusion, contingent sum fusion (with z-normalization of scores [13]), and log likelihood ratio fusion. We compared the results with unimodal (fingerprint or face only) and always-sum fusion of these biometrics. Figure 5 shows results of such comparisons, where Table 3 shows their FNMR_T and cost at $\text{FMR}_T = 0.001$. Cost equation for decision and contingent sum is as in equation 3, while cost for always fusing face and finger is as in equation 13. In all cases $c_1 = 7.5$, $c_2 = 15$ and $c_3 = 900$ seconds. The results are quite interesting. Adding face will noticeably improve performance of a single fingerprint verification system; it

Table 3. Multi-modal contingent fusion. FNMR_T and cost are computed at FMR_T = 0.001.

fusion scheme	FNMR _T	cost(seconds)
unimodal - finger	0.0205	25.9
unimodal - face	0.18	181.8
contingent decision	0.0042	11.8
contingent sum	0.0065	13.8
always sum	0.0067	33.5

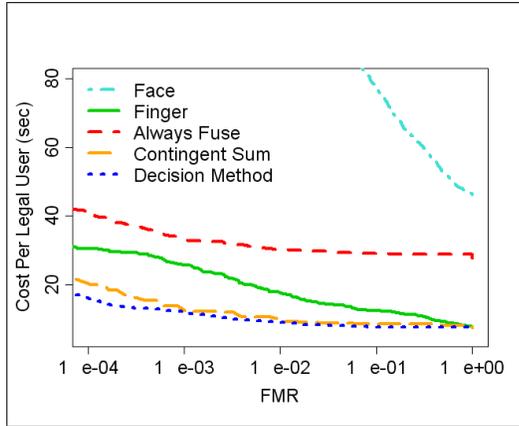


Figure 5. Multi-modal (finger and face) contingent fusion

decreases both FNMR_T and cost at a fixed FMR_T. While error rates of contingent and always-sum fusing face and finger are comparable, contingent sum’s cost is about a third of always-sum fusion of face and finger. Multi-modal contingent decision gives the lowest error rate and lowest cost of all multi-modal fusion scenarios studied.

3. Discussion

We find that the use of contingent fusion significantly reduces the average throughput time per user and FNMR_T at a fixed false match rate. All three contingent fusion schemes (decision, sum fuse, and likelihood ratio) consistently provide better performance, in terms of both error rate and cost, than always fusing two-biometric method. Among all, decision fusion, while the simplest, seems to work almost as well as the other two. The performance discrepancy becomes more apparent for lower false accept rates, but never gets worse than being within 2 seconds of the best fusion strategy. Likelihood ratio fusion improves throughput time slightly, but increases the complexity of the system due to requiring the prior knowledge of score distributions, which if feasible, might not be robust. The advantage realized by likelihood fusion stems from using the additional quality information to weigh the two scores differently during fusion.

Multi-instance and multi-modal fusion are more effective than multi-algorithm. An effective and successful multi-algorithm fusion depends on the appropriate choice of matchers and their operational thresholds.

In the single finger case (unimodal), users are either passed, or sent to secondary inspection. For relatively low false accept rates, the single finger method has a relatively high processing time due to the increase in the number of users that would be referred to secondary inspection. The crossover in cost that occurs between the single and double finger strategies is where the advantage of only having to process a single finger equals the advantage of superior identification gained by using two fingers.

More generally, using contingent fusion instead of the two-finger strategy can reduce about 5 seconds while still maintaining the same error rates. Better fusion strategies can improve performance, however the use of contingent fusion provides more substantial gains. For every fusion scheme, the cost appears to decrease about linearly with the log of the false match rate.

4. Conclusion

We solved the optimization problem of equation 7 empirically by iterating over all possible thresholds of test₁ and test₂ (τ_1 and τ_2) and plotting the minimum cost $C(\tau_1, \tau_2)$ for various FMR of the fused system, when scores are fused at either decision or score level. We studied multi-instance, multi-algorithm and multi-modal fusion scenarios and compared unimodal, fusing two biometrics at all times, and our proposed contingent fusion for each scenario. We found that contingent fusion reduces the cost (measured as the elapsed time from capturing an image until a decision is made to accept or reject the claim of identity) and error rate (FNMR at a fixed FMR) compared with unimodal and always fusing two biometrics. Fusing two instances of a biometric (*i.e.* impressions of right and left index) or two traits of biometric (*i.e.* impression of right index and face) works better than fusing two algorithms, probably because they are less correlated. Furthermore, contingent decision level fusion works as well as contingent score level fusion and is simpler to implement.

References

- [1] J. Kittler, M. Hatef, R. P. Duin, and J. G. Matas, "On Combining Classifiers", In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, pages 226–239, 1998. **1, 3**
- [2] Y. Wang, T. Tan, and A. K. Jain, "Combining Face and Iris Biometrics for Identity Verification", In *Proceedings of the Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 805–813, 2003. **1**
- [3] K. A. Toh, X. Jiang, and W. Y. Yau, "Exploiting Global and Local Decision for Multimodal Biometric Verification", In

IEEE Transaction on Signal Processing, pages 3059–3072, 2004. 1

- [4] A. Ross, K. Nandakumar, and A. K. Jain, "Handbook of Multibiometrics", *Springer*, 2006. 1
- [5] P. Grother, M. McCabe, C. Watson, M. Indovina, W. Salamon, P. Flanagan, E. Tabassi, E. Newton, and C. Wilson, "MINEX: Performance and Interoperability of the INCITS 378 Fingerprint Template", *National Institute of Standards and Technology*, NISTIR 7296 edition, March 2006. 1, 2
- [6] J. Daugman, "Biometric Decision Landscapes", Technical Report No. TR482, University of Cambridge Computer Laboratory, January 2000 1
- [7] L. Wein and M. Baveja "Using Fingerprint Image Quality to Improve the Identification Performance of the U.S. Visitor and Immigrant Status Indicator Technology Program", In Proceedings of the National Academy of Sciences of the United States of America, May 2005 2
- [8] E. Tabassi, C. L. Wilson, and C. I. Watson, "NIST fingerprint image quality", *National Institute of Standards and Technology*, NISTIR 7151 edition, 2004. 2, 3
- [9] E. Tabassi and C. L. Wilson, "A novel approach to fingerprint image quality", In *IEEE International Conference on Image Processing 2005*, Genoa, Italy, September 2005. 3, 5
- [10] A. Jain, K. Nadakumar, A. Ross "Score Normalization in Multimodal Biometric Systems", In *Pattern Recognition*, Vol. 38, No 12, pp.2270-2285, December 2005. 3
- [11] A. Jain, Y. Chen, K. Nadakumar, S. Dass "Score Normalization in Multimodal Biometric Systems", NIST Biometric Quality Workshop, March 2006, <http://www.itl.nist.gov/iad/894.03/quality/workshop/presentations.html>. 4
- [12] C. Wilson, P. Grother, R. Chandramouli "Biometric Data Specification for Personal Identity Verification", <http://csrc.nist.gov/publications/nistpubs/800-76/sp800-76.pdf>, February 2006. 4
- [13] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain "Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems" In *IEEE Transactions on Pattern Analysis and Machine Intelligence* pp.450-455, 2005. 6