NORTHROP GRUMMAN

Lantern Accuracy Analysis Engineering Report

30 November 2006

Programme Name: IDENT1
Contract No.: 3333
Document Set: Project Management
Document ID: CCN014R2-020.1-1.0

Prepared for:
Police Information Technology Organisation (PITO)

Prepared by:
Northrop Grumman Information Technology

Approval Signature

Lantern Accuracy Analysis Engineering Report

30 November 2006

Programme Name: IDENT1
Contract No.: 3333
Document Set: Project Management
Document ID: CCN014R2-020.1-1.0

Prepared for:
Police Information Technology Organisation (PITO)

Prepared by:
Northrop Grumman Information Technology

Prepared by:
Lee Whitney

Approved by:
Jerry Munizza
Development Engineering Manager

Approved by:
Daryl Salmons
Systems Engineering Manager

Approved by:
Rod Forry
Programme Manager

Approved by:
Mark Kieffer
System Assurance Manager

1

Revision History

Date	Document ID	Revision/Change Description
30 November 2006	CCN014R2-020.1-1.0	Initial release for PITO comments and/or acceptance

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Overview

This engineering report provides a summary of the Operating Point Analysis (OPA) performed by Northrop Grumman to predict and optimise the accuracy of Lantern searches during the Lantern Phase 2 Pilot.

This accuracy is based on test searches with known truth; that is, the correct search results are known in advance. This allows any misses or false rejects (Type 1 errors) and any false accepts (Type 2 errors) to be identified and used to calculate error statistics on a set of test searches. Because the truth is uncertain when searching the national fingerprint database with prints of unknown subjects, it is impossible to measure error rates obtained in actual operation. Therefore, this analysis reports a projection of the accuracy expected in operation. Actual accuracy depends on variable factors such as image quality, which is strongly influenced by capture technique. This test attempted to replicate authentic operational conditions and methods to the extent practicable so the results could be considered realistic.

The analysis also provides guidance for the setting of fingerprint matcher score thresholds, which determine whether a subject is reported as a match or not. Lantern search results are not verified by a fingerprint specialist, so the decision logic governing reporting based on matcher scores is crucial to the operational accuracy of the method. Determination of score thresholds to define the best operating point is a matter of trading the probability of reporting a true match against probability of incorrectly reporting a false match.

1.2 Background

Lantern captures images of two index fingerprints, and transmits them to a central site with a search request. There, they are encoded to extract features of the fingerprints that are compared against the national Ten Print collection. High speed matchers perform thousands of matches per second and return a match score that depends on how closely the features of the search print align with the file print.

1.2.1 Search Print Collection

Fingerprints for use in OPA searches were captured from police officers using actual Lantern Mobile Fingerprint Readers (MFRs). MFRs were issued to PITO personnel who had made prior arrangement to capture the prints of volunteer officers from two of the police forces. The volunteer subjects were from the Dyfed Powys and Essex forces, which were to be equipped with Lantern operational capability later in the Pilot. Their fingerprint images were recorded along with identifying collar numbers and names for correlation with stored elimination prints.

1.2.2 Elimination File Prints

In order to obtain search results from which accuracy can be determined, it is necessary for the search prints to have known true mates in the file to be searched. Elimination prints had been previously captured from the volunteers and other police officers for use in crime scene investigations. These prints are used to eliminate from consideration any marks that may have been inadvertently left by police personnel in the course of securing the scene or gathering evidence. When a mark is found at a crime scene, it is searched against the elimination prints and discounted as evidence if it matches that of an officer.

A copy of the Dyfed Powys and Essex Police Eliminations feature files were added to a copy of the Unified Collection and stored on servers and RAID located in the UK. This comprised the features file against which the search data collected via MFR was matched.

1.2.3 Preparation of the Collected Data

It was determined that approximately 500 or more sets of prints would be needed in order to have adequate confidence in the results, especially if the targeted high accuracy values were achieved. For example, the target false accept rate (FAR) of 0.5 percent allows just one in every 200 searches, or less than three total in 500 searches. This makes for rather large granularity in the results, but it was a trade-off against pragmatic considerations about the difficulty in collecting usable prints whose mates were on file.

Just under 500 subjects' prints were actually collected, but some had to be omitted from the analysis because their mates could not be found in the elimination file as expected. A careful visual examination of each search print against all possible mates was undertaken. The following possible errors were considered:

- Mate not as indicated by collar number, and the correct mate could not be found
- Mate not as indicated by collar number, and the correct mate could be found
- Mate is as indicated by collar number, but fingers are interchanged
- Mate is as indicated by collar number, but a duplicate mate is also present.

These could have been caused either by errors in recording their true identities during collection, or by errors/omissions in the eliminations data. When a subject's search prints did not match the file prints specified as truth, the subject was deleted from the data set. The remaining total of 452 subjects were confirmed as usable data and declared to be ground truth.

1.2.4 Previous Assessment

Under CCN009, a preliminary operating point assessment was performed and reported on 10 October 2005 (Doc. ID UIS037-21.2e-1.0). The assessment was performed using the matcher that existed at the time, and established accuracy targets against which to compare later assessments. The accuracy achieved in the previous assessment was:

- [redacted] at [redacted] FAR
- [redacted] at [redacted] FAR.

This matcher was not tailored for Lantern searches, where two plain index finger images are searched against the corresponding fingers from the database of rolled Ten Prints. Instead, they were as originally designed for searching marks against rolled Ten Prints. SAGEM subsequently developed modified algorithms for the matchers to optimise their performance with Lantern searches. The analysis reported herein found improved accuracy results from the new matcher algorithms.

1.3 Report Objectives

This document is submitted in satisfaction of two document delivery requirements defined in CCN014R2:

- Engineering Report: Operating Point Recommendation (matcher score threshold settings obtained from the OPA, recommended and implemented for Lantern operation)
- Engineering Report: Accuracy Analysis Report ("…define expected accuracy SLAs for application in an operational system…based on an analysis of tests carried out against data obtained from individuals known to be present on the IDENT1 database").

Because both these topics are based on the same analysis of the same data and done at the same time, a rationale exists for combining both into a single document. PITO concurred with this approach on 16 November 2006, and this document is the result.

No further testing to gather additional data for analysis is planned during the Pilot because cutover has occurred, putting the system into an operational mode. In operational mode, the fingerprint features file to be searched is automatically kept in synchronism with that of IDENT1 and the Police National Computer (PNC). Because of this, the elimination files are no longer present in the search population. It also precludes further seeding with individuals known to be present on the IDENT1 database to force matches of known truth.[1]

The content of this report necessarily has some technical complexity in the areas of statistical analysis, biometrics, and fingerprint identification system architecture in general; and Lantern and IDENT1 in particular. It is written for a target audience of non-specialists, but some background in those fields would be helpful.

[1] There is one exception to this. To facilitate training in October 2006, PITO made a special arrangement to have the PNC temporarily add one subject, the trainer, whose fingerprints were then entered into the IDENT1 and Lantern databases. When demonstrating the MFR, the trainer uses his own fingerprints for searches. When they match against this subject as expected, the results are returned with annotation from the PNC that the match is a special test subject rather than a criminal hit.

1

2. Reference Documents

Document	Document ID	Author	Date Issued
Lantern Operating Point Assessment	UIS037-21.2e-1.0	Northrop Grumman	10 October 2005
Contract Change Note 014, Revision 2; LANTERN Pilot Phase 2 – Implementation Phase	CCN014R2	Northrop Grumman	5 May 2006

2

3. Analysis

Matcher accuracy, as defined by probability of match (PM) and false accept rate (FAR), is dependent on two variables:

- Separation between matcher scores of true mate prints and non-mate prints
- Setting of score threshold for reporting a match.

This is illustrated graphically in Figure 3-1. The non-mates tend to score lower than the true mates. Unfortunately, the probability distributions of mate and non-mate scores can overlap, creating an area of ambiguity where errors occur. The less overlap, the lower the error rates. The threshold defines the score below which a match is not reported. The score where the threshold is set is a trade-off between false rejects (misses) and false accepts (hits). Set too low, it accepts and wrongly reports more non-mates; set too high, it rejects more true mates.

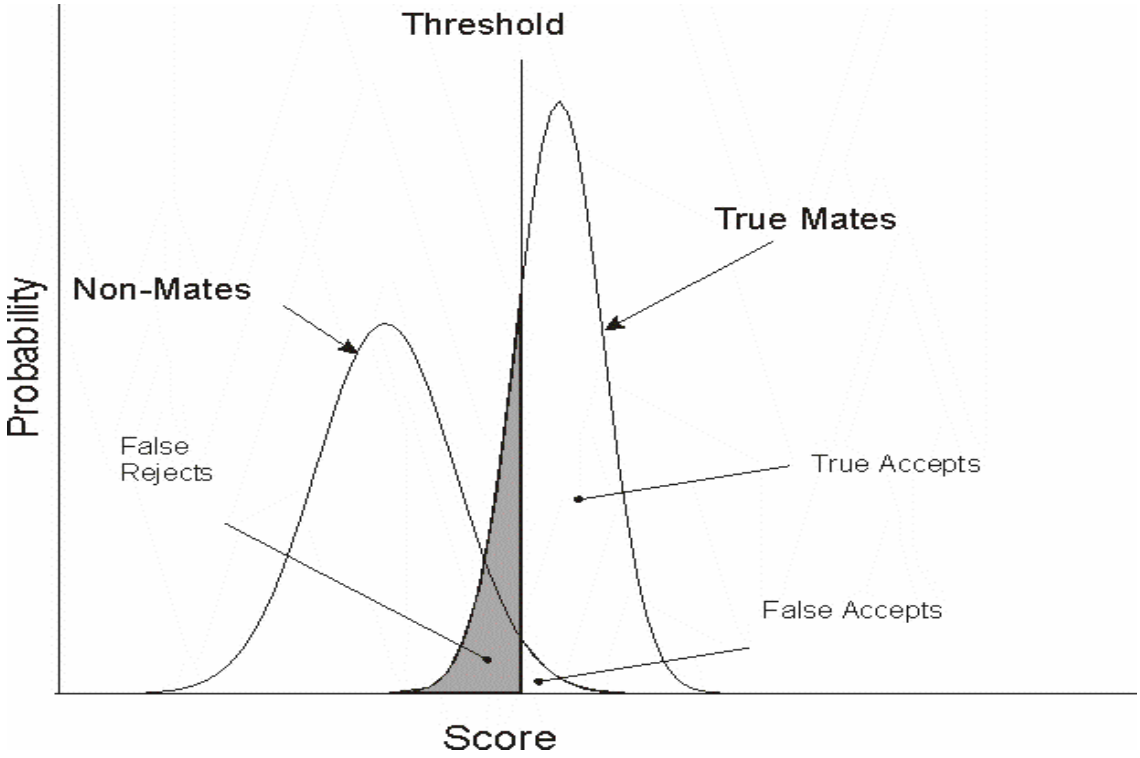


Figure 3-1 Illustrative Matcher Score Distribution

For any given set of scores obtained by matching fingerprints against their true mates and a background of non-mates, once the threshold is set the accuracy is established. Matching against non-mates creates background conditions that allow the false accept rate to be measured.

The goals of this analysis are to:

- Set the threshold for Lantern operation
- Determine the resulting accuracy for the test data set.

Given a test data set of fingerprints with realistic image qualities and a realistic volume of background fingerprints, the resulting accuracy is a reasonably good predictor of the accuracy to be expected in operation.

3.1 Initial Results

As reported at the Lantern Test Readiness Review (TRR) on 2 October 2006, initial values for the accuracy were determined early in the analysis. They were subject to refinement as the processing of the data was continuing, but gave a preliminary approximation of the expected accuracy.

The main points of the interim report on the Operating Point Analysis were the following:

- Interim results met minimum matcher accuracy requirements enabling testing and further analysis to proceed
- Matcher accuracy requirements were defined in CCN014 Pilot 287 as, "…that achieved by the Lantern Demonstrator (As reported in the "Lantern Operating Point Assessment" issued by Northrop Grumman on 10 October 2005.)" That document reported achieving PM = [redacted] at [redacted] FAR
- Actual measured interim accuracy = [redacted] PM at [redacted] FAR
- Analysis was continuing to further improve matcher accuracy
- SAGEM was consulting on threshold settings
- Report was to be delivered documenting final results (this document).

The initial results showed that the SAGEM MicroMatcher with Lantern modifications was an improvement over the previous version even without further adjustments.

3.2 SAGEM Analysis of the Data

Raw and summary data was provided to SAGEM for their analysis. The raw data consisted of combined 2-finger scores from matching each set of search prints against each set of file prints, both true mates and non-mates. It also included truth as defined by collar numbers collected with the fingerprints and keyed to the elimination file prints.

Summary data was generated as follows. Using the linear sum of match scores of the two index fingers, a cumulative distribution function (CDF) was computed for mates and non-mates. This result is shown on a per search basis in Figure 3-2. The separation of the two curves is a measure of the ability of a matcher to distinguish true mates from non-mates.

Figure 3-2 Mate and Non-mate Cumulative Distribution Functions by Combined Score

Figure 3-2 is scaled to focus on the lowest scores—those under 200. These account for over 0.9 or 90% of the non-mates, but only an extremely small percentage of mates. The mate curve does not quite reach zero as expected because of occasional true mates that did not appear among the 60 highest scoring respondents of each search that were considered for the graph. Therefore, their combined score was unreported and was treated as zero in the analysis. In the region below the 60^{th} respondent on a search, there is little reason for interest because these cases are simply considered misses.

Figure 3-2 data was then converted into a Receiver Operating Curve (ROC) plotting PM versus FAR over a range of threshold values. Figure 3-3 shows the results using two methods of conversion. The "All Data" method bases the FAR on the expected number of false respondents for each search, converted to the probability that a search has one or more errors. The second plot, labelled "Top One Only" uses only the top scoring respondent, exploiting the strong tendency for the true mate to appear as the top respondent regardless of its score.

Examination of the extreme right of the plot shows that with a PM of [redacted], the estimated FAR is about [redacted] for the "All Data" case and between [redacted] and [redacted] for the "Top One Only" case. There is not enough data to improve the precision of this estimate, and that FAR is too high to be useful for Lantern operation anyway. It does show that in this region, the "Top One Only" method gives better results than "All Data."

A strict reading of the graph shows a PM of [redacted] at a FAR of [redacted], but this also is in a region of the plot that suffers from insufficient data. It takes just three false hits to exceed a FAR of 0.005. This corresponds to a combined score threshold of [redacted]. At a FAR just above [redacted] the PM rises to [redacted] with a good sample size. The combined score threshold at this point is [redacted]

SAGEM continued with a more detailed analysis of specific false hits that appeared questionable. Careful visual inspection of the search and file prints of two supposed false hits were in fact true hits against a subject whose prints were actually in the file multiple times but had not been so indicated in the definition of truth. These false hits were corrected, lowering the FAR rate substantially from [redacted] to [redacted] with the same [redacted]PM.

3.3 Establishment of Operating Points

The operating points are defined by thresholds for combined score set as externally controlled variables in the SAGEM MicroMatcher software. This software resides in a number of Knowledge-Based Matchers (KBMs) each executing multiple simultaneous threads, 48 total threads in the case of Lantern. Two thresholds are required for Lantern: high and medium confidence. The high confidence threshold defines the score above which a reported match is considered most certain (low FAR). The medium confidence threshold defines the score above which a match is more likely to be reported, but less certain than a high confidence match, and there may be up to three medium confidence respondents reported.

Table 3-1 summarises the operating points that were established by SAGEM's analysis and recommended for implementation by SAGEM, and their expected accuracy. It also states the logic by which the thresholds are applied in determining whether or not to report a respondent.

Table 3-1 Operating Points Summary

Confidence Level	Score Threshold	Reporting Rule	Accuracy (PM @ FAR)
High	[redacted]	Report only the highest scoring respondent at or above the high confidence threshold	[redacted]
Medium	[redacted]	Report up to 3 respondents with the highest scores at or above the medium confidence threshold provided that none are also at or above the high confidence threshold	[redacted]

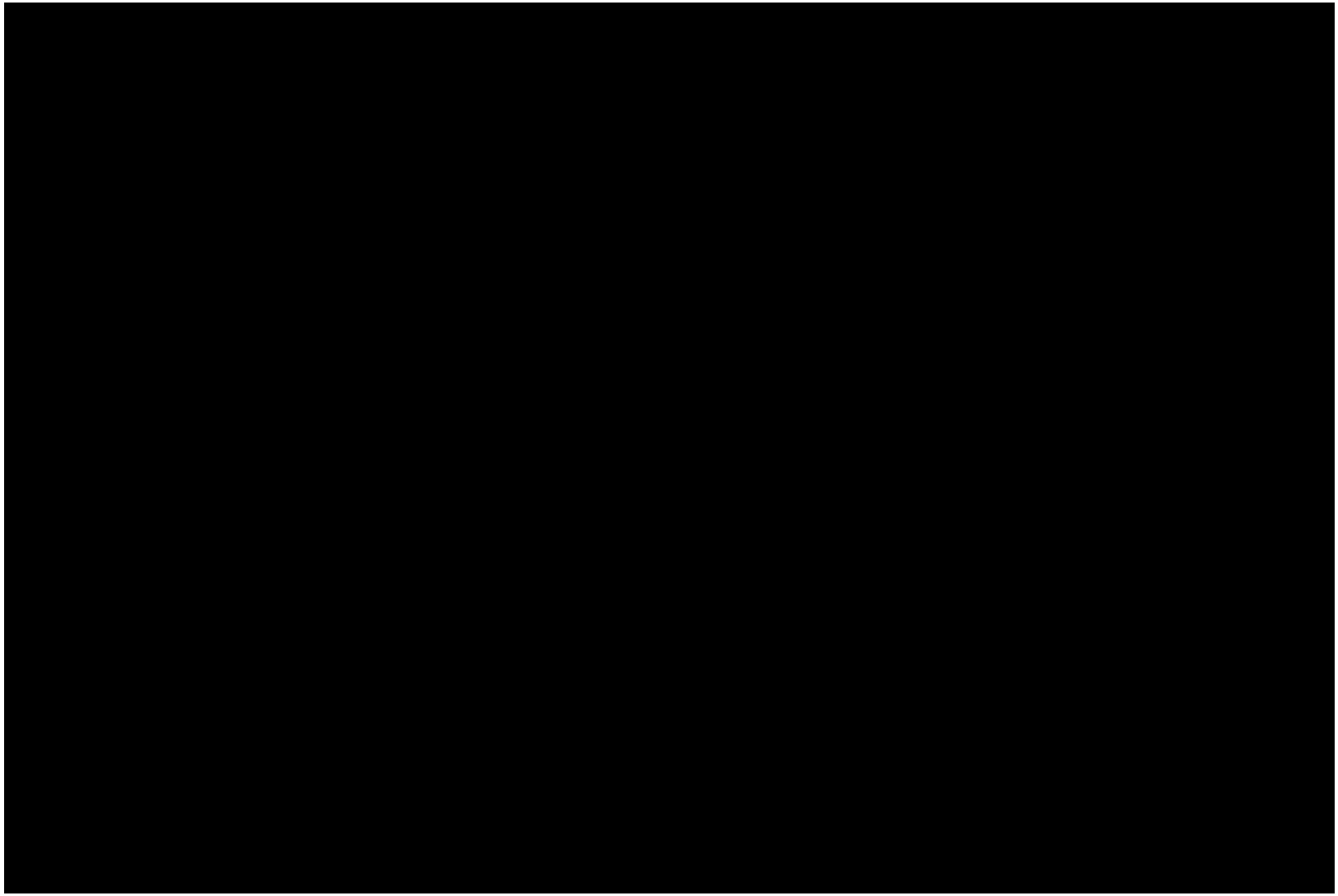
These operating points were implemented in all matchers on 11-October 2006 from [redacted]. The User Test immediately began observing the expected results on its seeded subject searches, i.e., predominantly high-confidence respondents.

3.4 Observations on the Data Quality

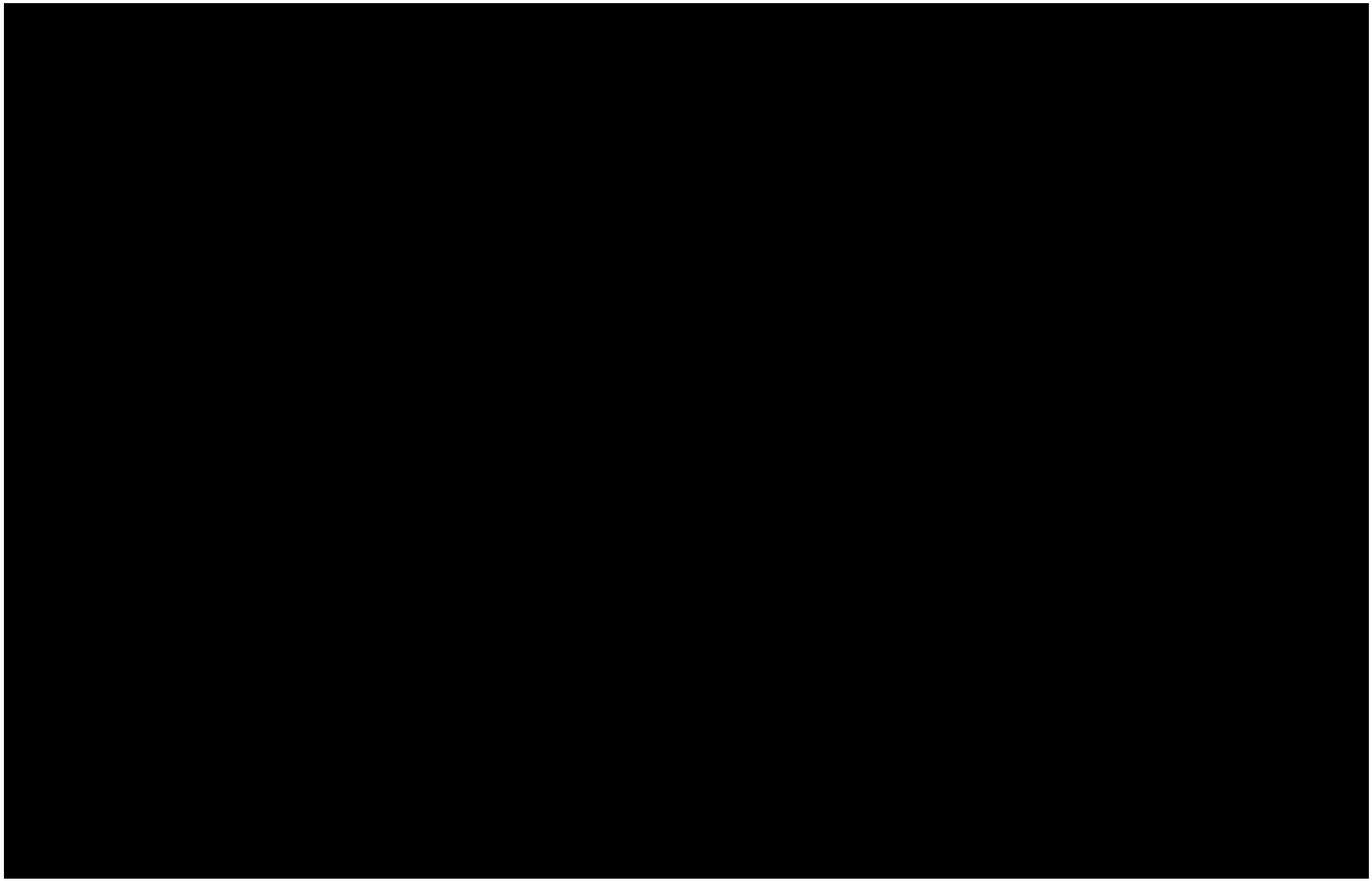
SAGEM's analysis of the match results from the Operating Point Analysis drew attention to an important fact. The accuracy was limited by some searches having an insufficient number of minutiae that were shared between the search print and the file print. Analysis of the misses showed that they predominantly shared fewer minutiae than hits, and not surprisingly they scored substantially lower.

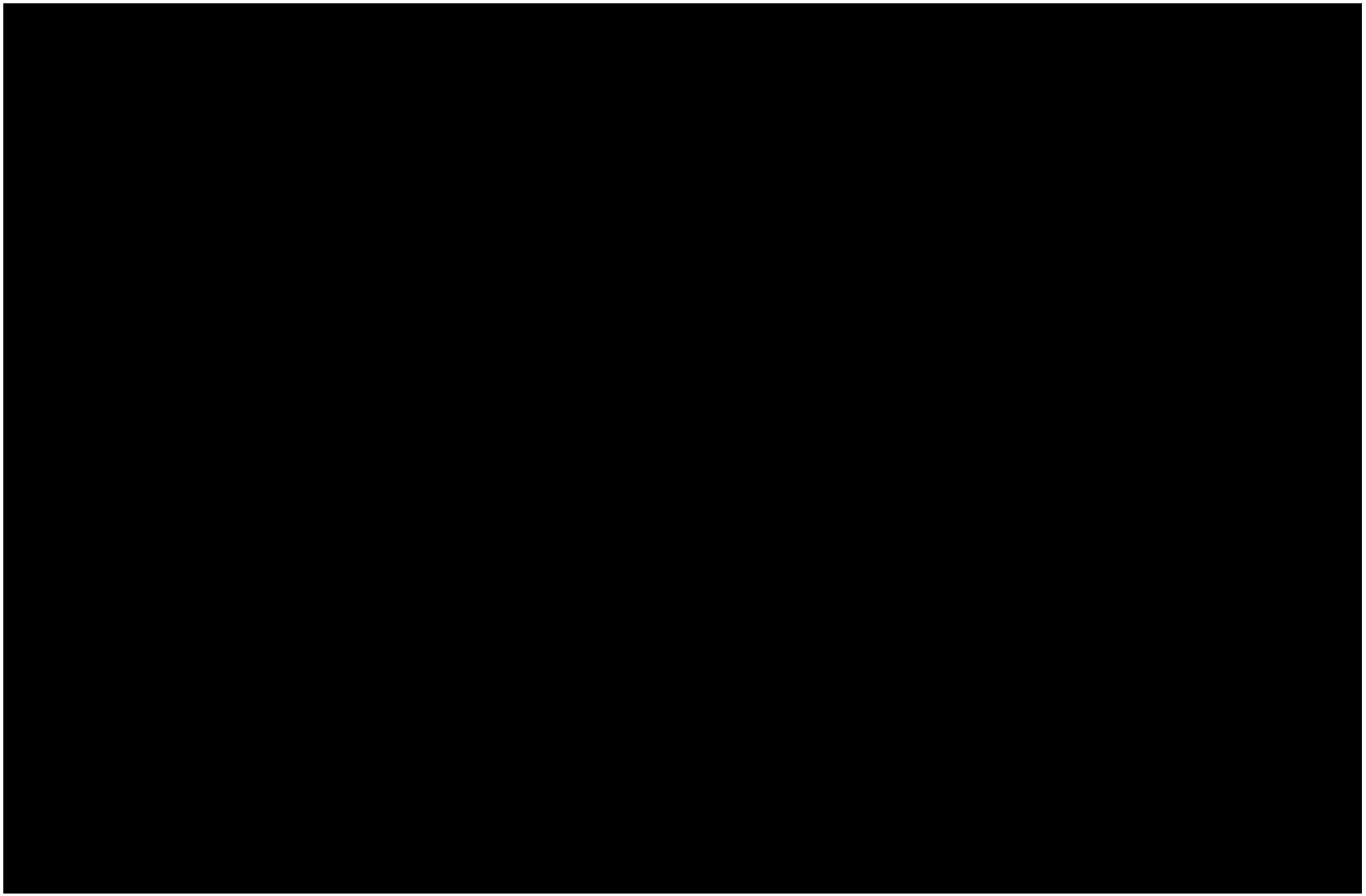
The scatter plot shown in Figure 3-4 was created by SAGEM[2] to show the number of minutiae (nb minutiae) of all search (enquiry) prints and their top-ranked mates as a function of their resulting matcher score. Most search print images had significantly fewer minutiae than the file prints they matched. This is as expected because they are plain impressions of a smaller area in contrast to rolled ink impressions or Livescans. However, the plot shows that the lowest scoring search prints shared only a fraction of their minutiae with their file matches. In other words, those few minutiae that were captured for the search prints were not found to be in common with many minutiae on the file mates.

[2] Figures 3-4, 3-5, and 3-6 are used without alteration by permission of SAGEM. French terms are "nomber" for "number" and "cumulated" for "cumulative." "Enquiry" refers to search prints.



Figures 3-5 and 3-6 compare the number of shared minutiae and the resulting scores respectively for the Lantern OPA data and a reference database of slaps and rolls taken by SAGEM from an operational site. In both shared minutia count and match scores, the Lantern data is visibly inferior to the reference data.





Visual examination of the images showed that this was often caused by improper finger positioning on the platen of the MFR when their images were captured. Fingers were typically positioned too low, inclined at an angle to the platen, oriented off image-vertical, or rotated about the finger axis. Emphasis on fingerprint capture technique has been increased in training in an attempt to overcome this in operation.

4. Conclusions

4.1 Recommended Operating Points

The analysis shows that excellent results are achieved with the thresholds set as indicated by the analysis ([redacted] for high confidence and [redacted] for medium). It is recommended that these operating points be retained for operational use.

4.2 Accuracy Expectations

The accuracies achieved by the matchers using the recommended operating points on the test data set described herein exceed the requirement, which was based on the previous assessment. This suggests that the Lantern matchers perform better on Lantern searches than their predecessors. The actual accuracy results were as follows:

- High confidence: PM=[redacted], FAR=[redacted]
- Medium confidence: PM=[redacted], FAR=[redacted]

Lowering the high threshold might yield marginally more hits with high confidence, but at the expense of increased FAR.

Medium adds mostly false hits. So, lowering either threshold would probably just add more false hits without a large increase in true hits.

The statistical certainty of these results is limited by the number of searches (less than 500). If it had been much larger, the FAR on high confidence results might not have been one false hit. Regardless, the search population was over 6.5 million, which is realistic for the national fingerprint database; and that is the number that drives the FAR in a direct proportion. So, the numbers are as realistic as can be for the size of the available test data set.

Accuracy can be maximised through fingerprint capture techniques that results in the highest possible image quality on the search prints. Training for use of the MFR should continue to emphasise the importance of using best capture operating technique.