Automated Fingerprint Identification System (AFIS) Benchmarking Using the National Institute of Standards and Technology (NIST) Special Database 4

Rajiv Khanna and Weicheng Shen
The MITRE Corporation
7525 Colshire Drive
McLean, Virginia 22102

Abstract – The recent growth in the number of products that use fingerprint matching technology has spawned a need for appropriate evaluation methods. We propose a set of benchmarks based on NIST Special Database 4 to support the evaluation of fingerprint matching technology. The approach used to develop the benchmark parameters and performance measurement methods are discussed. Benchmark parameters were established with members of the law enforcement community. Characteristics that typically affect fingerprint matching algorithm performance are observed and measured. Variation of matcher performance due to these characteristics, and others, indicates a need for standard data (NIST Special Database 4) when comparing algorithm performance. Benchmark parameter measurements using the United Kingdom Home Office HO-39 fingerprint matching algorithm and a Sun SPARC 10 are provided. Analysis of benchmark results provides insight into matcher performance. An example demonstrates how the benchmarks may be used.

INTRODUCTION

Automated Fingerprint Identification Systems (AFIS) are used in law enforcement and security applications to identify individuals based on their fingerprints. The law enforcement application is typically a database search mechanism. The security application is typically a template matching problem. There are, however, applications that require combinations of these approaches. A pattern matching algorithm is typically at the heart of these systems. Pattern matching is used to compare two different fingerprint images to determine if they represent fingerprint impressions from the same individual.

A number of new AFIS products and changes to existing systems has spawned a need for appropriate evaluation methods. New AFIS products range from databases on personal computers to large-scale hardware pattern matchers coupled with mainframe computer databases. In addition, changes to systems such as the incorporation of livescan technology and data compression may affect overall system performance.

We propose benchmark parameters and associated evaluation methods to help AFIS users and developers measure system performance. The proposed benchmark will enable AFIS users and developers to make informed decisions regarding AFIS acquisition and development. The benchmark consists of a well defined set of benchmark parameters, a qualified test dataset, and a method for measuring the parameters. Universal adoption of standard data, parameters, and methods will allow for performance comparison among competing alternatives and note actual changes in system performance.

AFIS BENCHMARK DEVELOPMENT

Benchmark Testbed

We developed an Identification Technology Testbed in support of benchmark development and AFIS test and evaluation. The testbed contains a Sun SPARCstation model 10/54 with 128 MB RAM and 10 GB magnetic storage. The United Kingdom Home Office HO-39 fingerprint pattern matching algorithm and several additional algorithms are available to support fingerprint processing. There are over 75,000 fingerprint images in the testbed database. The testbed is linked to other MITRE laboratories and computers via an ethernet LAN.

Benchmark Parameters

Table 1 contains a list of benchmark parameters and definitions appropriate to the law enforcement community. We identified the parameters and presented them to law enforcement officials at local, state, and Federal levels. We also presented the parameters to several members of the United Kingdom Home Office. The parameter definitions were clarified and new parameters were added to the set based on received comments.

Many of the benchmark parameters, while defined using terminology appropriate to law enforcement applications, correspond to security applications. General interpretation of the parameters for security applications is provided below:

- The type I error rate is typically defined as the number of times a system fails to find a match divided by the number of possible matches. Reliability is one minus the type I error rate. It is an overall effectiveness measure for the AFIS.
- Selectivity is used in estimating the amount of labor required to support a stated reliability in systems where the pattern matcher aids fingerprint examiners. For example, if a system produces a 250 candidates over 100 searches, its selectivity would be 2.5. Some fully automated systems may not require the selectivity parameter.
- The type II error rate is typically defined as the percentage of times a system falsely indicates a match. False Hit rate is equivalent to the type II error rate.
- Consolidation Efficiency measures the capability of the AFIS to find duplicate entries in the database.
- Search Time measures the rate that the AFIS compares fingerprint features.
- Encode Time measures the rate that the AFIS extracts features from fingerprint images.
- Position Summary indicates where the mated fingerprint appears in sorted lists of matcher output.

All of these benchmark parameters will not necessarily apply to all systems. Some systems may be best described by a subset of these parameters

0-7803-1479-4/94/$4.00 ©1994 IEEE

Table 1. Benchmark Parameter Definitions

	Parameter	Definition
1	Reliability	Percentage of fingerprint mates that are correctly identified by the system.
2	Selectivity	Percentage of fingerprints that require review to support the measured reliability.
3	False Hits	Percentage of non-mating prints that are incorrectly identified by the system.
4	Consolidation Efficiency	Percentage of fingerprint mates that are identified as duplicates by the system.
5	Search Time	Average system time required to execute search.
6	Encoding Time	Average system time required to encode the benchmark data.
7	Position Summary	Percentage of mating fingerprints found in selected intervals of rank ordered lists of matcher output.

Benchmark Measurement

The process used to develop the benchmark is outlined in Figure 1. We are collecting data from a variety of sources and assembling a test database. A test dataset was selected from the database and processed using the HO-39 fingerprint matcher and other required algorithms. Baseline benchmark parameters were measured and are reported here.

NEED FOR STANDARD DATA

AFIS performance, like that for other pattern matching systems, is data dependent. Reporting performance parameter measurements on different input data may affect some or all of the performance parameters. The following analysis illustrates this point.

Comparison of Two Datasets

We received fingerprints collected by two different organizations: the California Department of Motor Vehicles (CALDMV) and NIST. The two datasets were collected and scanned in two different processes. The CALDMV dataset was collected using a Fingermatrix Mint-10 livescan device, printed using a laser printer, and reproduced using photocopy equipment. We then scanned the fingerprints using a Sharp JX-610 scanner at 600 dots per inch (dpi) and interpolated the images to 500 dpi for use with our fingerprint processing algorithms. The HO-39 fingerprint matcher was used to compute a comparison score between each pair of fingerprints in the dataset.

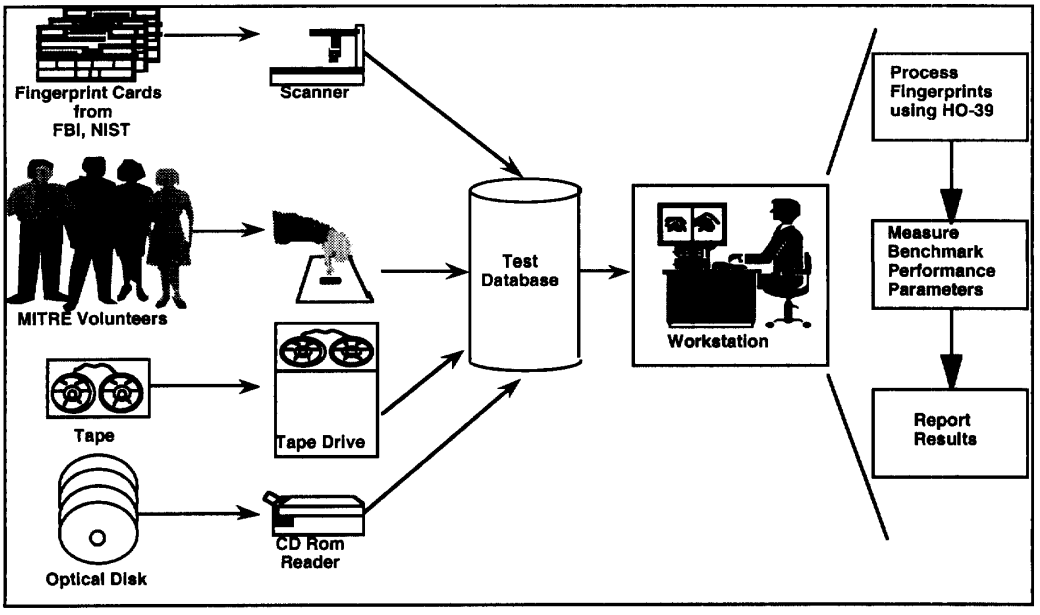

Figure 1. Benchmark Development Process

The NIST dataset was collected by rolling the finger on ink and then on an FBI standard fingerprint card. The fingerprints were then scanned by NIST at 500 dpi. We received the fingerprint images on optical disk. We processed the fingerprint images using the HO-39 in the same way as we processed the CALDMV dataset.

Analysis of Non-mated Fingerprints

The results of the this experiment indicate differences in the AFIS output for the two datasets. Figure 2 shows the histogram of AFIS output for non-mated fingerprints in the CALDMV dataset. The average score of the distribution shown in this histogram is 100. Figure 3 shows the corresponding histogram for non-mated fingerprints in the NIST dataset. The average score of this distribution is 150. There is a 50 point bias between the non-mate fingerprint scores that were generated by processing the CALDMV and NIST datasets.

Analysis of Mated Fingerprints

The analysis of AFIS output scores for mated fingerprints from the same two datasets is similar to the peviously described analysis. Figures 4 and 5 show histograms for CALDMV and NIST mated fingerprints, respectively. The average score for mated CALDMV fingerprints is 311. The average score for mated NIST fingerprints is 356. Again, the NIST fingerprints have higher average scores.

Analysis of Fingerprint Features

To further investigate the differences in average AFIS scores, we examined the average number of features per finger that were used by the AFIS. The CALDMV fingerprints averaged 56 features per fingerprint while the NIST fingerprints averaged 103 features per fingerprint. This may be one factor that contributed to the observed biases.

Standard Benchmark Data

The previous analyses indicate the need for standard data. Our AFIS produced different results on the two datasets. We would expect similar results using other AFIS. Development and acquisition decisions based on either one of the two datasets would be different. In order to make reliable and informed decisions, AFIS users and developers need to use standard data and evaluation methods. Our proposed benchmark was developed to help the fingerprint community fill this need.

The benchmark was developed using NIST Special Database 4. The fingerprint database is organized into five classes (or categories): whorl, left loop, right loop, arch, and tented arch. Each fingerprint class has 400 pairs of mated fingerprints, i.e., two rolled ink impressions for each finger. The fingerprints were collected from different fingers of a hand with no particular distribution.

NIST Special Database 4 was selected for a few fundamental reasons. It is widely available to the general public from NIST for a relatively low price ($250). NIST Special Database 4 contains fingerprints from each finger and contains fingerprints in each of 5 primary pattern classes. In addition, since it is available in scanned image form, the benchmark based upon NIST Special Database 4 measures post-scanning performance of the AFIS, eliminating biases caused by scanning errors.

BENCHMARK DEVELOPMENT AND RESULTS

Since many ten-print AFIS categorize input data by finger number prior to pattern matching, we partitioned NIST Special Database 4 into bins corresponding to each finger. We then used the HO-39 fingerprint matcher to compare each fingerprint in each finger category to every other finger in its category. The AFIS decision logic thresholds were set to the minimum value that yielded no False Hits, corresponding to an operating condition similar to that in some law enforcement applications. Benchmark parameters were measured using the resulting AFIS scores.

Tables 2 and 3 contain baseline benchmark parameters for the right and left hands respectively. We observed fairly uniform performance between the two hands. We also observed that the performance on the little finger is lower than that of the other fingers. This may be due to the smaller size of the little finger pattern area.

EXAMPLE OF BENCHMARK USE

To demonstrate application of the benchmark, we configured six different AFIS based on the HO-39 pattern matcher. The six AFIS are configured using three different types of decision logic. The first decision logic uses only the score to make match/no-match decisions. The second uses score and differences between scores. The third decision logic uses scores to simply rank order possible matches with the highest ranking fingerprint tested. In addition, three of the six AFIS are configured to be symmetric, while the remaining three are not.

Figure 6 illustrates the comparison of two benchmark parameters: Average Search Time and Reliability. The term "Enhanced" in the system descriptor indicates that symmetry was imposed on the system. The version number denotes the decision as described previously. The parameters are scaled for ease of review.

Best reliability appears in the AFIS denoted "Enhanced HO-39, V. 3." This AFIS is symmetric and uses scores only to rank order possible matches. But this AFIS also requires the most average time to conduct the search. "HO-39 V. 3" has the next best matcher reliability that performs at about 95 percent of the previous case. But this matcher requires only half the average search time. This matcher would provide the most "bang for the buck." In other words, it has a fairly high reliability but operates much faster.

CONCLUSIONS

We have proposed benchmarks and evaluation methods for AFIS. The benchmarks are based on standard data available from NIST: NIST Special Database 4. The benchmarks provide an open standard to compare fingerprint pattern matching systems. Since the data is available in image form, the benchmark does not measure scanning performance. Hence, additional testing needs to be used to evaluate scanner performance, and there is much literature and several test patterns available from NIST and IEEE.

This benchmark serves as a first step towards measuring AFIS operational performance by providing a test that measures image processing and pattern matching performance. We demonstrated that the benchmarks measure

performance improvements for some key properties of fingerprint pattern matching systems.

We demonstrated how the benchmark may be used to evaluate fingerprint pattern matching systems. These tests may be used to evaluate fingerprint matching algorithms at various phases: development, test, and acquisition.

We propose that developers and users of fingerprint identification systems apply our benchmarks and report system performance using these standard methods to benefit the fingerprint pattern matching community.

REFERENCES

[1] American National Standards Institute, Automated Fingerprint Identification Systems — Benchmark Tests of Relative Performance, 16 February 1988.

[2] J. A. W. Easthope, *Benchmarking of AFR Systems*, United Kingdom Home Office, 9 September 1993.

[3] B. J. O'Riordan and J. A. W. Easthope, *Benchmarking of the Home Office AFR Algorithms Print to Print Searching, Phase II*, United Kingdom Home Office.

[4] R. Khanna and W. Shen, *Proposed AFIS Benchmarking Using NIST Special Database 4*, presented at the International Association of Identification — Chesapeake Bay Division Annual Spring Conference Presentation, Richmond Virginia, 12 March 1994.

[5] R. Khanna, D. E. Lake and C. G. Bates, *The Identification Technology Laboratory*, presented at the Biometric Authentication Consortium Presentation, Albuquerque New Mexico, 25 October 1993.

[6] R. Khanna, D. E. Lake and C. G. Bates, *Overview of MITRE's Identification Technology Laboratory*, presented at the High Performance Computers and Communications Conference Presentation, Huntington WV, 19 October 1993.

[7] R. Khanna, D. E. Lake and C. G. Bates, *Evaluation Methods for Identification Systems*, presented at the SPIE Advanced Imaging and Pattern Recognition Conference, Washington DC, 13 October 1993.

[8] R. Khanna, D. E. Lake and C. G. Bates, *AFIS Benchmarking with an Identification Technology Testbed*, presented at the Criminal Justice Technology Symposium , Gaithersburg MD, 9 September 1993.

[9] C. I. Watson and C. L. Wilson, *NIST Special Database 4, Fingerprint Database*, National Institute of Standards and Technology, 17 March 1992.

Table 2. Baseline Benchmark Parameters for Fingers of the Right Hand

	Right Thumb (1)	Right Index (2)	Right Middle (3)	Right Ring (4)	Right Little (5)
Number of Fingerprints	384	442	404	400	352
Reliability (Percent)	75.52	71.00	77.23	78.00	52.84
Selectivity (per Search)	0.76	0.71	0.77	0.78	0.53
False Hits	0.00	0.00	0.00	0.00	0.00
Average Encoding Time (sec)	6.74	6.14	6.39	6.26	5.98
Average Search Time (sec)	506.37	429.17	486.62	467.49	360.10

Table 3. Baseline Benchmark Parameters for Fingers of the Left Hand

	Left Thumb (6)	Left Index (7)	Left Middle (8)	Left Ring (9)	Left Little (10)
Number of Fingerprints	396	438	432	396	356
Reliability (Percent)	72.22	73.52	75.93	70.20	53.93
Selectivity (per Search)	0.72	0.74	0.76	0.70	0.54
False Hits	0.00	0.00	0.00	0.00	0.00
Average Encoding Time (sec)	6.66	6.16	6.34	6.17	5.9
Average Search Time (sec)	538.99	511.38	503.86	460.22	342.06

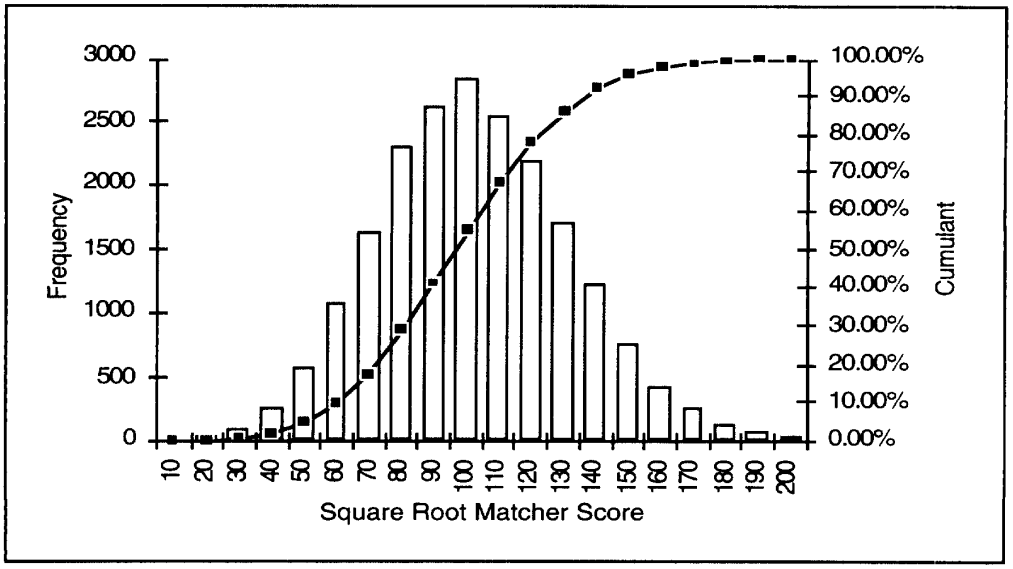

Figure 2. Histogram of AFIS Scores for Non-mated CALDMV Fingerprints

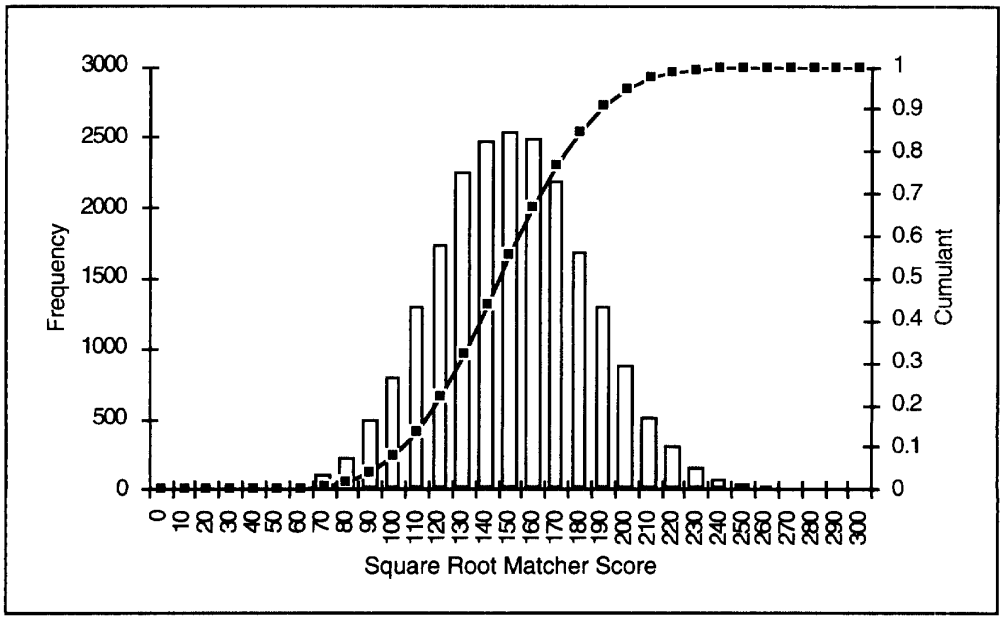

Figure 3. Histogram of AFIS Scores for Non-mated NIST Fingerprints

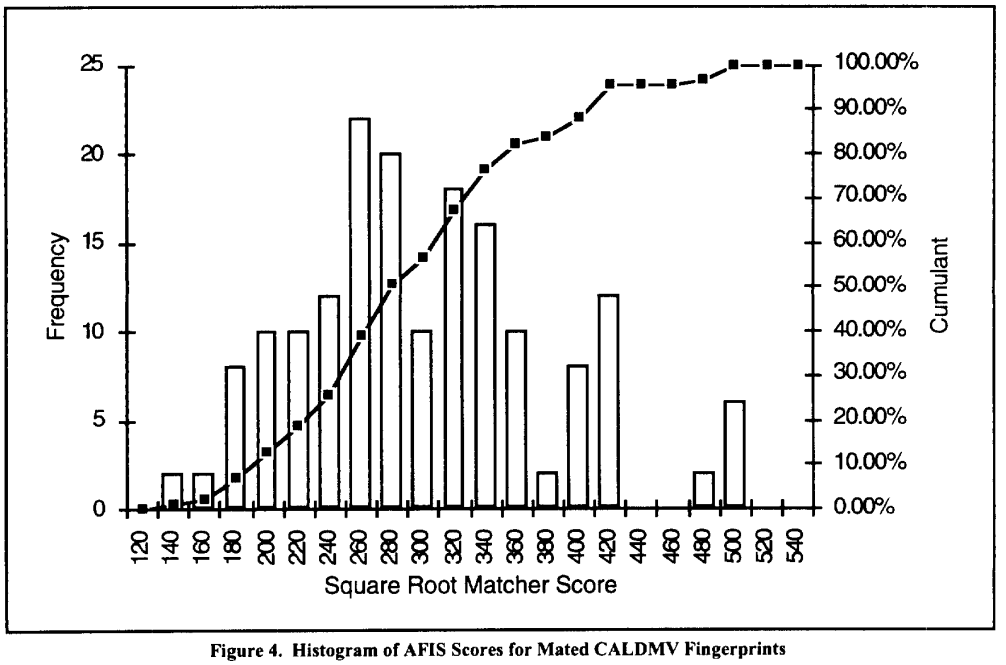

Figure 4. Histogram of AFIS Scores for Mated CALDMV Fingerprints

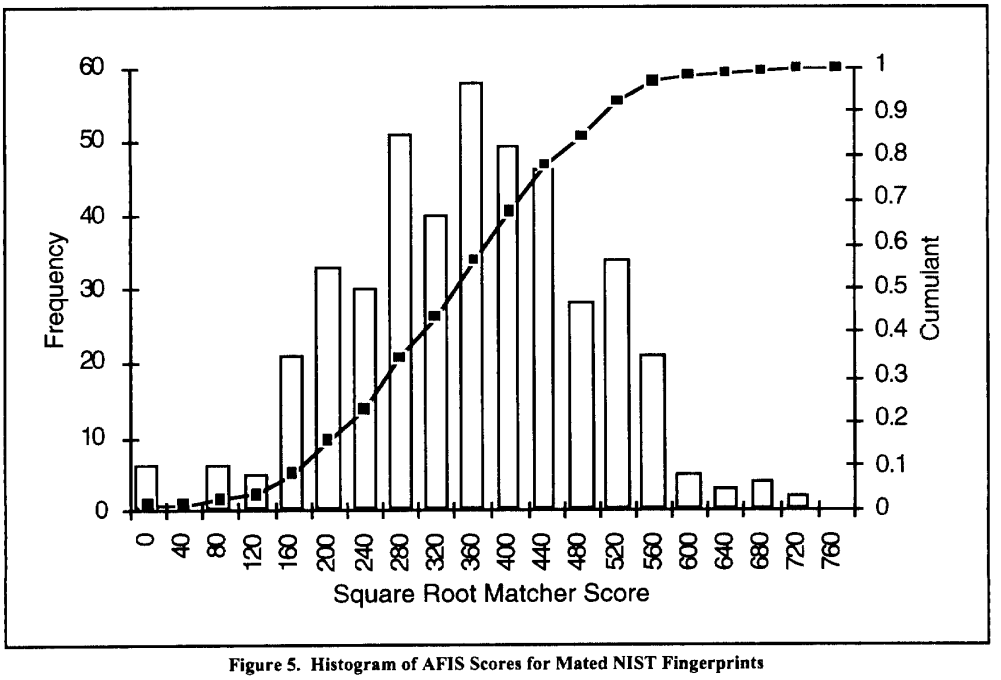

Figure 5. Histogram of AFIS Scores for Mated NIST Fingerprints

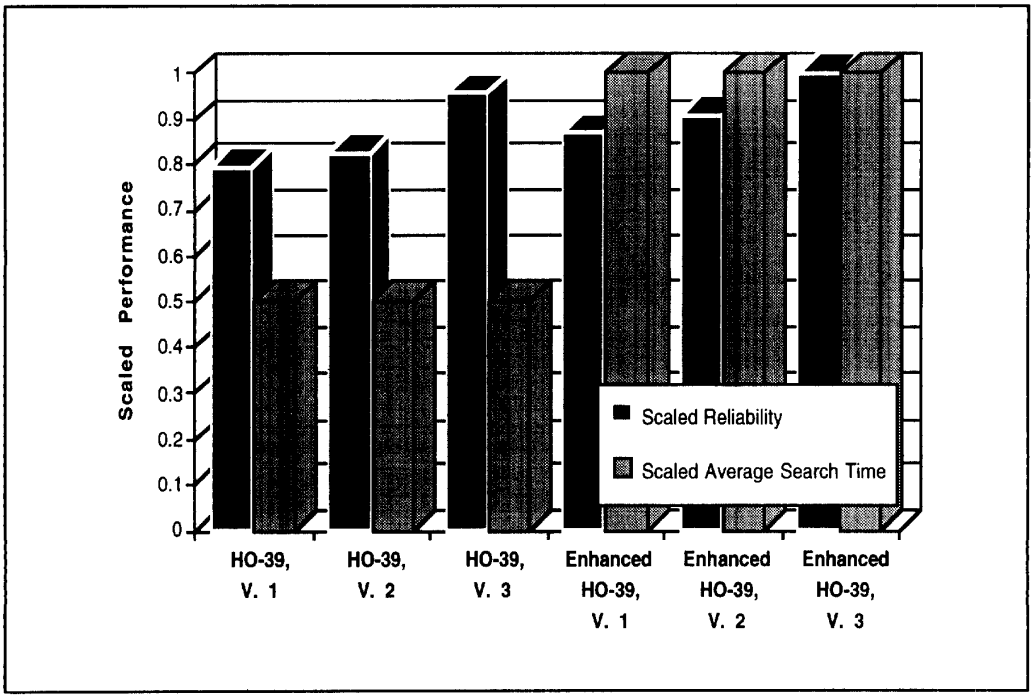

Figure 6. Comparison of Two Benchmark Parameters for Six AFIS