Facial Recognition Vendor Test 2000

Executive Overview

December 1, 2000

Duane M. Blackburn[1]
Mike Bone[2]
P. Jonathon Phillips, Ph.D.[3]

Sponsored by:
DoD Counterdrug Technology Development Program Office
Defense Advanced Research Projects Agency
National Institute of Justice

[1] NAVSEA Dahlgren Division
17320 Dahlgren Road
Dahlgren, VA 22448
BlackburnDM@nswc.navy.mil

[2] NAVSEA Crane Division
300 Highway 361
Crane, IN 47522
Bone_Mike@crane.navy.mil

[3]Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203
JPhillips@darpa.mil

Acknowledgements

The sponsors of the Facial Recognition Vendor Test 2000 would like to thank the following individuals for their assistance throughout these evaluations:

Mr. Patrick Grother, National Institute of Standards and Technology, for providing the software tools and consultation that enabled the remaining authors to perform this evaluation.

Mr. Tom Coty and Mr. Chris Miles from the National Institute of Justice, and Mr. George Lukes from the Defense Advanced Research Projects Agency for their support throughout the FRVT 2000 evaluations and for reviewing this document.

Mrs. Jo Gann, Office of National Drug Control Policy, for her advice in general and for reviewing this document.

Mrs. Kim Shepard, BRTRC, for providing the evaluation web sites and the many changes that followed.

Mrs. Elaine Newton, RAND, for reviewing this document.

Executive Overview

1. Introduction

The biggest change in the facial recognition community since the completion of the FERET program has been the introduction of facial recognition products to the commercial market. Open market competitiveness has driven numerous technological advances in automated face recognition since the FERET program and significantly lowered system costs. Today there are dozens of facial recognition systems available that have the potential to meet performance requirements for numerous applications. But which of these systems best meet the performance requirements for given applications?

Repeated inquiries from numerous government agencies on the current state of facial recognition technology prompted the DoD Counterdrug Technology Development Program Office to establish a new set of evaluations. The Facial Recognition Vendor Test 2000 (FRVT 2000), was co-sponsored by the DoD Counterdrug Technology Development Program Office, the National Institute of Justice, and the Defense Advanced Research Projects Agency, and was administered in May-June 2000.

2. Goals of the FRVT 2000

The sponsors of the FRVT 2000 had two major goals for the evaluation. The first was a technical assessment of the capabilities of commercially available facial recognition systems. The sponsors wanted to know the strengths and weaknesses of each individual system, as well as obtain an understanding of the current state of the art for facial recognition.

The second goal of the evaluation was to educate the biometrics community and the general public on how to present and analyze results. The sponsors have seen vendors and would-be customers quoting outstanding performance specifications without understanding that these specifications are virtually useless without first knowing the details of the test that was used to produce the quoted results.

3. FRVT 2000 Evaluation Methodology

The FRVT 2000 was based on the evaluation methodology proposed in “An Introduction to Evaluating Biometric Systems,” by P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki in IEEE Computer, February 2000, pp. 56-63, 2000. This methodology proposes a three step evaluation protocol: a top-level “Technology Evaluation”, followed by a “Scenario Evaluation”, and finally an “Operational Evaluation.”

3.1 Recognition Performance Test (A Technology Evaluation)

The goal of a technology evaluation is to compare competing algorithms from a single technology, which in this case is facial recognition. Testing of all algorithms is done on a standardized database collected by a "universal" sensor and should be performed by an organization that will not see any benefit should one algorithm outperform the others. The use of a test set ensures that all participants see the same data. Someone with a need for facial recognition can look at the results from the images that most closely resemble their situation and can determine, to a reasonable extent, what results they should expect.

The operation of the Recognition Performance Test in the FRVT 2000 was very similar to the original FERET evaluations that were sponsored by the DoD Counterdrug Technology Development Program Office. Vendors were given 13,872 images and were asked to compare each image to all of the other images (over 192 million comparisons). This data was used to form experiments that will show how well the systems respond to numerous variables such as pose, lighting, and image compression level.

3.2 Product Usability Test (A Limited Example of a Scenario Evaluation)

A scenario evaluation is an evaluation of the complete facial recognition system, rather than the facial recognition algorithm only. The participating vendors were allowed to choose the components (such as camera, lighting, etc.) that they would normally recommend for this scenario. These components play a major role in the ability of a facial recognition system to successfully operate in a live environment. Therefore, it was imperative that these components, and their interactions, be evaluated as a system using live test subjects.

The Product Usability Test is an example of a limited scenario evaluation. A full scenario evaluation would have used significantly more test subjects and lasted a period of weeks, but it would have also been done on only one or two systems. The participating vendors were not paid to have their systems evaluated for the FRVT 2000 so it would have been unfair to ask each of them to spend their own money to support a multi-week evaluation. The scenario chosen for the FRVT 2000 Product Usability Test was access control.

The Product Usability Tests consisted of two timed tests. The timed tests were used to measure the response time of the overall system for two different operational scenario simulations: the Old Image Database Timed Test and the Enrollment Timed Test. Each of the timed tests was performed for both verification and identification and was performed once with overhead fluorescent lighting and again with the addition of back lighting.

4. How to Use This Report

The FRVT 2000 evaluations were not designed, and this report was not written, to be a "Buyer's Guide for Facial Recognition." Consequently, no one should blindly open this report to a particular graph or chart to find out which system is "best". Instead, the reader should study each graph and chart, the types of images used for each graph and chart, and the test method that was used to generate the graphs and charts and determine how each of them relate to the problem the reader is trying to solve. It is entirely possible that some of the experiments performed in both the Recognition Performance and Product Usability portions of this evaluation have no relation to the problem a particular reader is trying to solve and should be ignored. Once a reader has determined which image types and tests are applicable to their problem, they will be able to study the scientific data provided and determine which system to use in their own scenario and operational evaluations. The goal of this report is to provide an assessment of where the technology was in the May-June 2000 time frame. When considering face recognition technology to solve a specific problem, the results in this report should be used as one of many sources in designing an evaluation for your specific problem.

To understand some of the basic terms and concepts used in evaluating biometric systems, see the glossary located in Appendix N.