FAQ

Documents

Facial Recognition
Vendor Test 2000

Face Recognition
Vendor Test 2002

Evaluation
Methodology

Home

FERET

Introduction

The [Department of Defense (DoD) Counterdrug Technology Development Program Office](#) sponsored the Face Recognition Technology (FERET) program. The goal of the FERET program was to develop automatic face recognition capabilities that could be employed to assist security, intelligence, and law enforcement personnel in the performance of their duties. The program consisted of three major elements:

- Sponsoring research
- Collecting the FERET database
- Performing the FERET evaluations

The goal of the sponsored research was to develop face recognition algorithms. The FERET database was collected to support the sponsored research and the FERET evaluations. The FERET evaluations were performed to measure progress in algorithm development and identify future research directions.

The FERET program started in September of 1993, with Dr. P. Jonathon Phillips, Army Research Laboratory, Adelphi, Maryland, serving as technical agent. Initially, the FERET program consisted of three phases, each one year in length. The goals of the first phase were to establish the viability of automatic face recognition algorithms and to establish a performance baseline against which to measure future progress. The goals of phases 2 and 3 were to further develop face recognition technology. After the successful conclusion of phase 2, the DoD Counterdrug Technology Development Program Office initiated the FERET demonstration effort. The goals of this effort were to port FERET evaluated algorithms to real-time experimental/demonstration systems.

FERET-Sponsored Algorithm Development Research

The FERET program was initiated with a broad agency announcement (BAA). Twenty-four proposals were received and evaluated jointly by DoD and law enforcement personnel. The winning proposals were chosen based on their advanced ideas and differing approaches. Five algorithm development contracts were awarded. The organizations and principle investigators selected were:

- Massachusetts Institute of Technology (MIT), Alex Pentland
- Rutgers University, Joseph Wilder
- The Analytic Science Company (TASC), Gale Gordon
- University of Illinois at Chicago (UIC) and University of Illinois at Urbana-Champagne, Lewis Sadler and Thomas Huang
- University of Southern California (USC), Christoph von der Malsburg

For phase 2, MIT, TASC, and USC were selected to continue development of their algorithms. The MIT and USC teams continued work on developing face recognition algorithms from still images. The TASC effort extended their approach to developing an algorithm for recognizing faces from video. The emphasis of the TASC effort was to estimate the three-dimensional shape of a face from motion and then recognize a face based on its shape. Rutgers' Phase 2 effort compared and assessed the relative merits of long-wave infrared (thermal) and visible imagery for face recognition and detection. The results of this study were presented in the paper "Comparison of visible and infrared imagery for face recognition" by J. Wilder, P. J. Phillips, C. Jiang, and S. Wiener in Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pages 182-187, 1996. [Available soon].

The FERET Database

A standard database of face imagery was essential to the success of the FERET program, both to supply standard imagery to the algorithm developers and to supply a sufficient number of images to allow testing of these algorithms. Before the start of the FERET program, there was no way to accurately evaluate or compare facial recognition algorithms. Various researchers collected their own databases for the problems they were investigating. Most of the databases were small and consisted of images of less than 50 individuals. Notable exceptions were databases collected by three primary researchers:

- Alex Pentland of the Massachusetts Institute of Technology (MIT) assembled a database of ~7500 images that had been collected in a highly controlled environment with controlled illumination; all images had the eyes in a registered location, and all images were full frontal face views.
- Joseph Wilder of Rutgers University assembled a database of ~250 individuals collected under similarly controlled conditions.
- Christoph von der Malsburg of the University of Southern California (USC) and colleagues used a database of ~100 images that were of controlled size and illumination but did include some head rotation.

The FERET program set out to establish a large database of facial images that was gathered independently from the algorithm developers. Dr. Harry Wechsler at George Mason University was selected to direct the collection of this database. The database collection was a collaborative effort between Dr. Wechsler and Dr. Phillips. The images were collected in a semi-controlled environment. To maintain a degree of consistency throughout the database, the same physical setup was used in each photography session. Because the equipment had to be reassembled for each session, there was some minor variation in images collected on different dates.

The FERET database was collected in 15 sessions between August 1993 and July 1996. The database contains 1564 sets of images for a total of 14,126 images that includes 1199 individuals and 365 duplicate sets of images. A duplicate set is a second set of images of a person already in the database and was usually taken on a different day.

For some individuals, over two years had elapsed between their first and last sittings, with some subjects being photographed multiple times. This time lapse was important because it enabled researchers to study, for the first time, changes in a subject's appearance that occur over a year.

The FERET Evaluations

Before the FERET database was created, a large number of papers reported outstanding recognition results (usually >95% correct recognition) on limited-size databases (usually <50 individuals). Only a few of these algorithms reported results on images utilizing a common database, let alone met the desirable goal of being evaluated on a standard testing protocol that included separate training and testing sets. As a consequence, there was no method to make informed comparisons among various algorithms.

The FERET database made it possible for researchers to develop algorithms on a common database and to report results in the literature using this database. Results reported in the literature did not provide a direct comparison among algorithms because each researcher reported results using different assumptions, scoring methods, and images. The independently administered FERET evaluations allowed for a direct quantitative assessment of the relative strengths and weaknesses of different approaches.

More importantly, the FERET database and evaluations clarified the state of the art in face recognition and pointed out general directions for future research. The FERET evaluations allowed the computer vision community to assess overall strengths and weaknesses in the field, not only on the basis of the performance of an individual algorithm, but in addition on the aggregate performance of all algorithms tested. Through this type of assessment, the community learned in an unbiased and open manner of the important technical problems that needed to be addressed.

Three sets of evaluations were performed, with the last two evaluations being administered

multiple times. See the table below for dates and groups evaluated. Algorithms developed under FERET-funding were required to participate in the FERET evaluations. Other organizations were invited to participate in the FERET evaluations, but received no funding from the FERET program to do so. These organizations were Excalibur Corp, Michigan State University, Rockefeller University, Rutgers University (Sep96 evaluation only), and University of Maryland.

		Test date					
Version of test	Group	Aug 94	Mar 95	Nov 95	Aug 96	Sep 96	Mar 97
Fully automatic	MIT	*	*		*	*	
	Rockefeller U.			*			
	Rutgers	*					
	TASC	*					
	USC	*	*				*
Eye coordinates given	Baseline PCA	*					
	Excalibur					*	
	MSU					*	
	Rutgers					*	
	UMD					*	*

The first FERET evaluation took place in August 1994, the Aug94 evaluation, and was designed to measure performance on algorithms that could automatically locate, normalize, and identify faces from a database. The test consisted of three subtests, each with a different gallery and probe set. The first subtest examined the ability of algorithms to recognize faces from a gallery of 316 individuals. The second subtest was the false-alarm test, which measured how well an algorithm rejects faces not in the gallery. The third subtest baselined the effects of pose changes on performance.

The second FERET evaluation took place in March of 1995, the Mar95 evaluation. The goal was to measure progress since the initial FERET evaluation, and to evaluate these algorithms on larger galleries (817 individuals). An added emphasis of this evaluation was on probe sets that contained duplicate images, where a duplicate image was defined as an image of a person whose corresponding gallery image was taken on a different date.

The third, and final, FERET evaluations took place in September of 1996, referred to as the Sep96 FERET evaluation. For the Sept96 evaluation, we designed a new evaluation protocol which required algorithms to match a set of 3323 images against a set of 3816 images. Thus, algorithms had to perform approximately 12.6 million matches. The new protocol design allowed the determination of performance scores for multiple galleries and probe sets, and perform a more detailed performance analysis. Results were reported for the following cases: (1) the gallery and probe images of a person were taken on the same day under the same lighting conditions, (2) the gallery and probe images of a person were taken on different days, (3) the gallery and probe images of a person were taken over a year apart, and (4) the gallery and probe images of a person were taken on the same day, but with different lighting conditions. There were two versions of the September 1996 evaluation. The first tested partially automatic algorithms by providing the images with the coordinates of the center of the eyes. The second tested fully automatic algorithms by providing the images only.

Additionally, the designers of the evaluations implemented two face recognition algorithms (PCA and Correlation) to provide a baseline performance. To provide a greater understanding of face recognition algorithms, a detailed study of PCA-based recognition algorithms was performed and the results are in "Computational and Performance Aspects of PCA-based Face Recognition Algorithms" by H. Moon and P. J. Phillips, to appear in Perception, (NISTIR) [Available soon].

Summary

The DoD Counterdrug Technology Development Program Office began the Face Recognition (FERET) program in 1993 and sponsored it through its completion in 1998. Total funding for the program was in excess of $6.5 million.

The FERET program consisted of three major elements. First was sponsoring research that advanced facial recognition from theory to working laboratory algorithms. Many of the algorithms that took part in FERET form the foundation of today's commercial systems. Second was the collection and distribution of the FERET database, which contains 14,126 facial images of 1199 individuals. The DoD Counterdrug Technology Development Program Office still receives requests for access to the FERET database, which is currently maintained at the National Institute of Standards and Technology. The development portion of the FERET database has been distributed to over 100 groups outside the original FERET program. The final, and most recognized, part of the FERET program was the FERET evaluations that compared the abilities of facial recognition algorithms using the FERET database.

The test methods used in the FERET evaluations form the foundation of an overall biometric evaluation methodology described in "An Introduction to Evaluating Biometric Systems," by P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki in IEEE Computer, February, pp. 56-63, 2000. (Special issue on biometrics.) [Download here]. This evaluation methodology has been incorporated into the UK Biometrics Working Group in their "Best Practices in Testing Performance of Biometrics Devices" [Download here]. As clearly shown, the FERET program continues to have a profound effect on the facial recognition community today.

The FERET program was a highly successful effort that provided direction and credibility to the facial recognition community. We are just now beginning to uncover how important the program was during the infancy of facial recognition technology. As FERET nears the end of its transition from active program to a historical program, the DoD Counterdrug Technology Development Program takes great pride on the imprint it has left on the biometrics community, and even greater pride that the FERET ideals and evaluation methods are being used by current programs both inside the Program Office and by other Government agencies.

Information about obtaining a copy of the FERET database can be found in the NIST FERET Database page.