Face Recognition Vendor Test (FRVT) 2006 Protocol

10 October 2005

Introduction

The Face Recognition Vendor Test (FRVT) 2006 is the third in the FRVT series of evaluations. The two previous FRVTs were FRVT 2000 and FRVT 2002. The FRVT evaluations were preceded by the three FERET evaluations. FRVT 2006 is the same in spirit as the previous evaluations, an independent evaluation of face recognition algorithms on sequestered data. However, the overall structure of the FRVT 2006 is substantially different than in previous FRVT and FERET evaluations. The first major change is that FRVT 2006 will evaluate performance on four tasks: 1-1 matching, 1-many matching, all duplicate detection, and preprocessing. Reporting image quality is an option for 1-1 matching and 1-many matching. Similarity score normalization is also an option for 1-1 matching. The increase in the number of tasks and associated options represent an increase in the maturity and sophistication face recognition algorithms and systems.

The interface between the executables will be based on the Biometric Experimentation Environment (BEE). The BEE is being used as part of the Face Recognition Grand Challenge (FRGC) and the Iris Challenge Evaluation (ICE). The interface specifications for FRVT 2006 are provided in a separate document *Executable Calling Signatures for FRVT 2006.*

Academic groups were the primary participants in the FERET evaluations. Whereas, only companies participated in FRVT 2000 and 2002, despite FRVT 2002 being open to academia, research institutions, and companies. One of the goals of FRVT 2006 is to encourage academic groups to participate in the evaluation. To this end, FRVT 2006 will evaluate performance of both fully and partially automatic versions of algorithms and systems.

Academic research groups usually publish the details of their algorithms. Publication of the details of the algorithm combined with results from an independent evaluation, such as FRVT 2006, allows for a high level assessment of which algorithmic approaches are most promising. This assessment is usually not possible with commercial systems because the details of their system are not published. To provide a more systematic evaluation of published algorithms, FRVT 2006 will introduce the translucent evaluation tract. This tract is open to all, not just academic institutions, who publish the details of their algorithms or systems.

Finally, the last page of this document contains an open letter to potential FRVT 2006 participants. This letter briefly outlines the FRVT 2006 test protocol and data sets.

FRVT 2006 Tasks

Four tasks will be evaluated in FRVT 2006. Each participant will have to deliver a separate executable for each task. The four tasks are:

1. **1-1 Matching.** The basic biometric verification task is comparing two biometric samples and reporting a similarity. This task will measure performance when matching is based only on the two biometric samples being compared. Each similarity score between a target biometric sample t and query biometric sample q is independent of all other biometric samples in the target and query sets. The computation of image quality is an option for 1-1 matching. A single quality score is produced for each target and query image. The quality score for a biometric sample is an integer, with larger numbers indicating better quality. Another option available for this task is similarity score normalization. In similarity score normalization, the input is a set of similarity scores $s(G,Q)$, the executable returns a new set of similarity scores $s'(G,Q)$.

2. **1- Many Matching**. This task models the situation where subjects in a gallery are well known and the facial representation can be tuned to the gallery. The output is a complete similarity matrix of scores between a gallery G and a query set Q. Each set of similarity scores $s(G,q)$ are independent of all other query biometric samples. As with 1-1 matching, a complete similarity matrix must be computed. The computation of image quality is an option for 1-1 matching. A single quality score is produced for each target and query image. The quality score is an integer, with large numbers indicating better quality.

3. **Preprocessing**. This task will evaluate the effectiveness of preprocessing methods on 1-1 and 1-many matching. For FRVT 2006, the input will be a set of still images and the output will be a single still image in jpeg format. Preprocessed images will serve as input to 1-1 and 1-many matching experiments.

4. **All duplicate detection**. The goal of this task is to group and cluster all biometric samples of the same subject. The input is a biometric signature set of n biometric samples. A biometric sample could consist of multiple images (recordings). (For multiple recording biometric samples, all recordings will be from the same person.) The output is a label for each biometric sample with a number in the range $[1,...,n]$. All biometric samples with the same identity should have the same label. No biometric samples with different identities should have the same label. All labels will be integers in the range $[1,...,n]$.

General Rules

To maintain fairness, the Government will adhere to the following procedures for FRVT 2006. Information provided to any single Participant will also be provided to all Participants. The primary and preferred source of communications between Participants

and FRVT 2006 government personnel is the frequently asked question (FAQ) page (http://www.frvt.org/FRVT2006/FAQs.aspx). A government employee has been named as the FRVT Liaison. The FRVT Liaison will be the primary point of contact and will manage communications with Participants. Questions that are outside the purview of the FAQ can be submitted to the FRVT Liaison via email at frvt2006@nist.gov. All e-mails sent to the FRVT Liaison and germane to the FRVT 2006 will be posted on the FAQ page. This policy ensures all Participants are guaranteed equal access to information concerning FRVT 2006. Contact with the FRVT Liaison other than via the FAQ areas, or contact with someone other than the non-FRVT Liaison will only be allowed for extenuating circumstances (for example, delivery of login information for Participant Area or major issues with similarity files).

1. **Evaluation Result Output.** Complete results for all experiments must be outputted in the correct format for an executable. If complete results in the correct format are not provided, then results for an executable may not be released. For 1-1 and 1-many matching, and similarity score normalization, this means complete similarity matrices with floating point numbers for all entries. For image quality, this means an image quality for all biometric samples in a signature set. For preprocessing this means, an image for all biometric signatures. For all duplicate detection, this means a result for each biometric signature.

2. **Format.** The still image format for FRVT 2006 is jpeg. The 3D data will be from a Minolta Vivid sensor and a Qlonerator sensor. The format for the Minolta Vivid and Qlonerator sensors will be same as in the FRGC v1, v2 and supplemental datasets.

3. **Release of Evaluation Results.** Upon the completion of the evaluation, the Government will combine all results into a final report and other supplemental reports. The final report will contain, at a minimum, descriptive information concerning FRVT 2006, descriptions of each experiment, evaluation results, and each Participant's five-page system description document. A pre-release version of this report will be made available to Participants. Participants will be invited to provide comments which will be included as an appendix to the FRVT 2006 Evaluation Report.

More specific guidance concerning the report and Participant comments will be provided at a later date. Participants are reminded that they will not comment publicly or privately on the pre-release version of the FRVT 2006 Report until it has been released to the public. Participants should be aware that the FRVT 2006 final report and supplemental reports will attribute performance with the organizations/groups that participate.

4. **Additional Information.**   Any data obtained during these evaluations, as well as any documentation required by the Government from the participants, becomes the property of the Government. Participants will not possess a proprietary interest in the data and/or submitted documentation.

The Government is not bound or obligated to follow any recommendations of the Participant. The United States Government, or any individual agency, is not bound, nor is it obligated, in any way to give any special consideration to FRVT 2006 Participants on future contracts.

If a Participant decides to use results of these evaluations in any way for their own purpose, it must be accompanied by the following phrase: "Results shown from the Face Recognition Vendor Test 2006 do not constitute endorsement of any particular system by the Government." It must also be accompanied by an internet hyperlink to the final report on the FRVT 2006 web site.

Translucent Algorithm Evaluations

The details of the majority of algorithms developed by academics are published in the academic literature. From the details published in the literature and independent performance results it is possible to determine what are the most effective techniques in face recognition. Towards this end, FRVT 2006 will be offering a translucent tract. To participate in the translucent tract groups must provide a detailed description of their algorithm and permission for the inclusion of the description in FRVT 2006 reports, and posting on FRVT 2006 and NIST web pages. One of the goals of the translucent tract will be to evaluate the impact of different algorithm configurations and training sets on performance. To study the different algorithm configurations, participants will be able to have "control knobs" that can vary key design elements of algorithm parameters. Examples of different design configurations are distance metric or kernel in a classifier. Examples of different algorithm parameters are the number of features in a representation or the band-pass frequency for a preprocessing filter.

Participants in the translucent tract must submit both training and recognition modules. The accompanying write-up must describe all components in the algorithm sufficient so that a reasonable re-implementation of the algorithm can be performed. Components that must be described include, but are not limited to, preprocessing, feature generation, training, recognition, classifiers, and normalization procedures.

Details on applying to participate in the translucent tract will be included in the instructions for applying to participate in the FRVT 2006.

Combinations of Executables that can be Submitted

FRVT 2006 consists of four tasks, however, there are only certain combinations that participants can submit. Preprocessing modules can be submitted as with submission of any other modules. If a fully automatic version a module is submitted, then a partially automatic version of that module must also be submitted.

To submit a matching module (1-1 matching, 1-many matching, and all duplicates), a participant is required to submit a 1-1 matching module. A list of allowable combinations of matching modules is given in the table below.

1-1 Matching	1-many Matching	All Duplicates
*		
*	*	
*		*
*	*	*

Rules for Executables

This section governs the rules for the executables that are submitted to FRVT 2006.

Each FRVT 2006 participant will be given a naming convention for their executables. The naming convention will be provided later after completion of the registration procedure.

A separate executable will need to be submitted for each task.

Participants can submit multiple executables for a task. Examples of multiple executables are: fully automatic for still images, partially automatic for still images, and fully automatic for 3D images. If more than one executable is submitted, then an explanation needs to be provided for as for the reason for multiple executables.

Executables need to be able to be installed multiple times during FRVT 2006. We will erase disks and reinstall the operating system and executables to insure the evaluation protocols are being followed.

Executables will need to be able to run on a stand alone machine.

There will be no internet access during the FRVT 2006 evaluations. Executables will need to installed and executed without access to the internet.

Executables will need to able to run in either WindowsXP or Linux. The targeted versions of the operating systems are WindowsXP service pack 2 and Fedora 3. After a

participant has completed the registration process, details of the operating system, software, and hardware will be made available.

A FRVT 2006 participant will be provided with a directory structure for running their executable. There will be a separate directory structure for each executable. The name of the directory structure for each executable will be given during the registration process. Call this directory the home directory. The home directory will contain four subdirectories: bin, lib, temp, and output. These four directories will exist at the time an executable is installed. Participants are free to create subdirectory structures in these three directories.   All binaries and executable data files must be contained in a directory structure in the bin directory. Libraries can be placed in either the bin or lib directory structures. The executable itself must be in the bin directory. Data that an executable needs should be stored in the bin directory structure. If a subdirectory structure is needed, then it should be created at the time executable is installed. After the executable installation process is complete, the bin and lib directory structures cannot be modified. Note: the installation of an executable and its supporting data files and libraries could be as simple as unzipping or untarring a file.

Partial and temporary results can be stores in the temp directory structure. This should be viewed as a temporary structure. At the completion of an experimental run, the contents of the temp directory structure maybe erased. During a sequence of experimental runs, a biometric sample may be reused. To expedite processing, templates could be stored in the temp directory structure and reused. However, the code should be designed to handle the case were the template was erased. Partial results can also be stored in the temp directory.

Independence requirements

Parts of the FRVT 2006 will be run as a batch process, i.e., modules will be given a list of biometric samples and asked to process them. However, batch processing in FRVT 2006 is to expedite the evaluation of tasks that may not be inherently a batch process for potential applications of interest. To make sure the FRVT 2006 results generalize to these applications, a set of independence rules have been established. To avoid confusion, the independence of rules for each module/task will be explicitly stated. **As part of FRVT 2006, executables will be explicitly tested to make sure that the independence rules are obeyed. Failure to follow the independence rules will most likely result in a participant being disqualified from FRVT 2006.** Listed below are the independence rules for each module:

1. **1-1 matching**: The computation of a similarity score $s(t,q)$ is independent of other biometric samples in the both target and query sets. A simple test for independence, a similarity score $s(t,q)$ from a batch process must be yield the same results as when the target set only consist of the t, and the query set only consists of q.
   a. **Optional Image Quality:** The image quality rating $IQ(t)$ for target sample t, must be independent of all other biometric samples in the target signature set. A simple test for independence, $IQ(t)$ from a batch process must be the same as when the target set consists of a single target sample t.
   b. **Similarity score normalization**: The computation of the set of similarity scores $s(G,q)$ between a query image and a gallery G is independent of other biometric samples in the query set. A simple test for independence, a set of similarity $s(G,q)$ from a batch process must be the same as the results when the gallery sig set consists of G and the query set only consists of q.
2. **1-many matching**: The computation of the set of similarity scores $s(G,q)$ between a query image and a gallery G is independent of other biometric samples in the query set. A simple test for independence, a set of similarity $s(G,q)$ from a batch process must be the same as the results when the gallery sig set consists of G and the query set only consists of q.
   a. **Optional Image Quality:** The image quality rating $IQ(t)$ for target sample t, must be independent of all other biometric samples in the target signature set. A simple test for independence, $IQ(t)$ from a batch process must be the same as when the target set consists of a single target sample t.
3. **Preprocessing**: The preprocessing results of a biometric sample t, must be independent of all other biometric samples in the signature set. A simple test for independence, the results from preprocessing biometric sample t from a batch process must be the same as when the target set consists of a single target sample t.
4. **All duplicate detection**: There are no independence requirements on processing a target set.

Training

FRVT 2006 does not dictate or require specific training or tuning sets. FRVT 2006 participants are free to train and tune their algorithm/system on the data set(s) of their choice. Many FRVT 2006 participants are likely to be Face Recognition Grand Challenge (FRGC) participants. As part of FRGC, the data was partitioned into training and validation sets. This partition does not apply to FRVT 2006, FRVT 2006 participants are free to train on all the data provided as part of FRGC. **With the exception of the translucent tract, all algorithms and systems must be submitted completely trained and tuned.**

Experiment Information Provided

The experimental protocol will provide information about each experiment. The information that will be provided is:

- Whether the experiment is partially or fully automatic.
- Whether the target or gallery samples are still or 3D.
- Whether the query samples are still or 3D.
- Whether the target or gallery images are all controlled or all uncontrolled or mixed.
- Whether the query images are all controlled or all uncontrolled or mixed.
- The minimum and maximum number of recordings per biometric sample for the target set or gallery.
- The minimum and maximum number of recordings per biometric sample for the query set.

This information will be provided in the parameter file that accompanies each experiment.

Notation:

The definition of target, query, gallery, and probe sets is based on [Face Recognition Vendor Test 2002 Performance Metrics](), by P.J. Grother, R.J. Micheals and P. J. Phillips. Proceedings for the 4th International Conference on Audio Visual Based Person Authentication, 2003 (http://www.frvt.org/DLs/Avbpa_2003_evaluation_metrics.pdf). Below is a list of notations used in the document:

- t target biometric sample
- q query biometric sample
- T target set
- Q query set
- G gallery
- $s(t,q)$ similarity score between target biometric t and query biometric sample q.
- $s(G,q)$ similarity scores between Gallery G and query biometric sample q.
- $s(G,Q)$ the complete set of similarity scores between Gallery G and query set Q.

- $IQ(t)$ image quality measure for biometric sample t.

Open letter to potential FRVT 2006 participants

A large number of potential FRVT 2006 participants took part in the Face Recognition Grand Challenge (FRGC). The groups that participated in FRGC worked hard and made significant advances in face recognition. Because FRVT 2006 is an independent evaluation there is a different set of preparations that need to be completed for algorithms and systems to successfully compete. This open letter contains some comments on the differences and is a guide to help groups prepare for FRVT 2006.

As I have mentioned previously, FRVT 2006 is not test on performance of FRGC data or on data collected at Notre Dame. Rather it is a test of face recognition algorithms and the ability of algorithms and systems to generalize to new sets of facial images. It is advisable to validate the performance on your algorithm on different data sets. Examples of other data sets include FERET and Banca.

The FRGC data set was divided into training and validation partitions. Since FRVT 2006 will not measure performance on the FRGC data set, FRVT 2006 participants are free to use the entire FRGC corpus of data to train their algorithm/system. In fact, FRVT 2006 participants can use any data to train and tune their algorithm.

FRVT 2006 executables need to be able to perform recognition on large data sets of images that have not been previous processed. This means the executables need to be robust and able to recover from errors and successfully complete runs on large data sets such as found in FRGC v2. This means that there are no memory leaks that could cause an algorithm to crash or run very slowly. Algorithms/systems need to handle cases where modules fail. For example, what happens if an eye location module fails to find an eye? What happens when a template is not successfully generated? What happens if the ground for the eye coordinates is faulty? What happens if there is garbage in an image? What happens if there is a bad image or your routine fails to read an image? In all cases you should be able to produce a complete similarity matrix. If an image is good, then your code definitely needs to gracefully handle the failure. If the problem is on the testing end, then we will rectify it.

A reasonable effort will be made to determine if the cause of failure to complete an experiment is an error at the testing site (i.e., a bad file). The purpose of the log file is to assist the FRVT 2006 testers in diagnosing these errors. If the log file is not present, it will be hard to diagnosis the cause of an error. One of the goals is to identify and correct errors.

There are a large number of options available with FRVT 2006. Preparing for FRVT 2006 may seem daunting. However, preparing for and participating in the basic evaluation, 1-1 matching, should be within reach of most research groups.

One of the goals of this open letter is provide guidance to potential FRVT 2006 participants as they prepare. In the past, we have found that such guidance does help, and insures that participants can successfully complete the evaluation, avoid mistakes, and maximize their performance.