REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* *8/16/2004*	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* 6/1/2003-5/31/2004

4. TITLE AND SUBTITLE Face Detection and Recognition	**5a. CONTRACT NUMBER** DAAD05-03-C-0045
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Anil K. Jain	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Michigan State University Department of Computer Science & Engineering 3115 Engineering Building East Lansing, Michigan 48824	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) TSWG SCOS Program Manager 1111 Jefferson Davis Highway, Suite 116 Arlington, VA 22202	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This report describes research efforts towards developing algorithms for a robust face recognition system in order to overcome many of the limitations found in existing two-dimensional facial recognition systems. Specifically, this report addresses the problem of detecting faces in color images in the presence of various lighting conditions and complex backgrounds as well as recognizing faces under variations in pose, lighting, and expression. The report is organized in two main parts: (i) Face detection and (ii) face recognition. A near real-time face detection system has been developed that uses a skin-tone color model and facial features. For face recognition, we have developed four independent solutions: (i) Evidence accumulation for 2D face recognition, (ii) demographic information extraction from 2D facial images, (iii) 3D model enhanced 2D face recognition with small number of training samples, and (iv) 3D face recognition.

15. SUBJECT TERMS

Face recognition, face detection, 3D face model, feature extraction, matching

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Anil K. Jain
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**		7	**19b. TELEPHONE NUMBER** *(include area code)* 517-355-9282

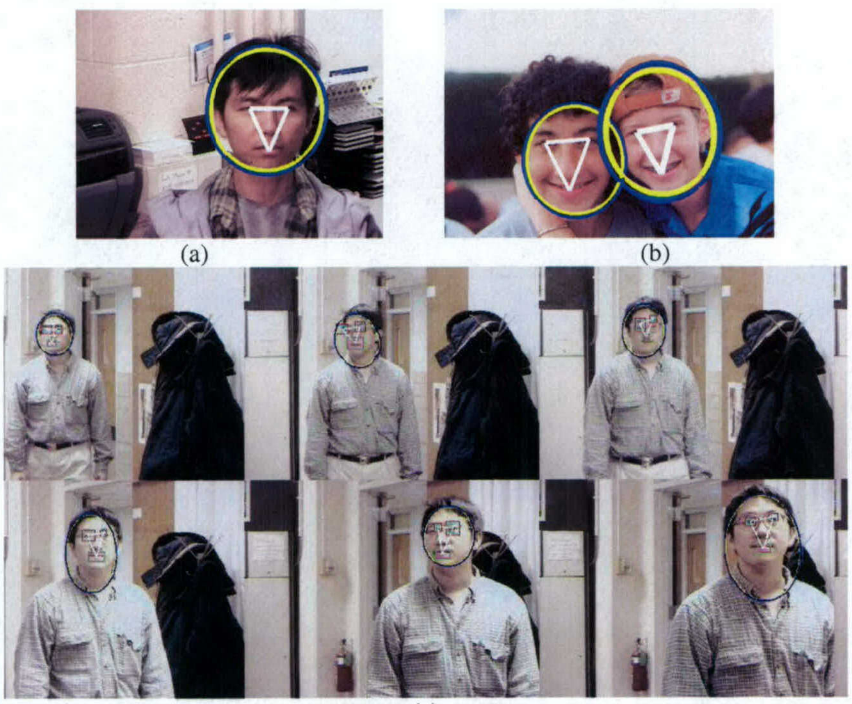
Face Detection and Recognition Final Report
BAA# DAAD05-03-T-0023
Mission Area: Surveillance Collection and Operations Support
Anil K. Jain
Michigan State University
August 16, 2004

I. Introduction

The goal of this effort is to develop new algorithms for a robust pose-invariant face recognition that overcome many of the limitations found in existing facial recognition systems. Specifically, we are interested in addressing the problem of detecting faces in color images in the presence of various lighting conditions and complex backgrounds as well as recognizing faces under variations in pose, lighting, and expression. This work is separated into two major components (i) Face detection and (ii) Face recognition. Specific tasks include developing modules for face detection, pose estimation, face modeling, face matching, and a user interface.

II. Face detection

We have developed a robust, near real-time face detection system from color images using a skin-tone color model and facial features. Major facial features are located automatically and color bias is corrected by a lighting compensation technique that automatically estimates the reference white pixels. This technique overcomes the difficulty of detecting the low-luma and high-luma skin tones by applying a nonlinear transform to the color space. We have also developed a robust face detection module to extract faces from cluttered backgrounds in still images (See Figures 1 a and b) The system is easily extended to work with video image sequences (See Figure 1c). The proposed system not only detects the face, but also locates important facial features, such as eyes and mouth. These features are crucial to the performance of the face recognition. See [1] for algorithm details. The total computation cost to both face detection and feature localization for a 640x480 image is less than 10 seconds on a 2.7G Hz CPU. It varies due to the complexity of the image.



(a) (b) (c)

Figure 1. Face detection and facial feature localization. (a) and (b) are results for static images. (c) demonstrates the result for a video sequence where a person is walking into the room.

III. Face recognition

The problem of face recognition in a general situation (arbitrary pose, lighting and facial expression) is a very difficult problem. In this project we have successfully investigated a verity of different approaches for achieving our goals in face recognition. We have developed four independent solutions to face recognition systems that investigate different aspects of our project goals:

1. Evidence accumulation for 2D face recognition
2. Demographic information extraction from 2D facial images
3. 3D-model enhanced 2D face recognition with few training samples
4. 3D face recognition

The first approach is a robust extension of existing standard (appearance based) face recognition methodology because it only uses 2D images for representation. The second approach investigates methods of indexing a large database of face images. Successful indexing allows the test images to be binned into groups that significantly reduce the number of comparisons that need to be made for face recognition. Our third approach extends 2D face recognition by using a more robust 3D model of the face to account for variations in expression. Our fourth approach uses a 3D scanner for both model building and acquiring test scans. Table 1 shows four combinations of scenarios where these different types of information (2D and 3D) could potentially be used to augment the identification process.

Currently the most common approach to face recognition uses a database (template) of 2D information to recognize 2D test images (upper left box in Table 1). In the first and second approaches we combine various successful approaches to 2D face recognition. Even this approach does not compensate for lighting and pose changes. However, 3D information inherently makes a face recognition system more robust to pose and expression variation. Approach 3 attempts to store face information as a generic 3D model of the face and then mach this model to 2D images (lower left box in Table 1). This approach is better because it does not require any special hardware for acquiring the face image. However, because we have access to a full 3D scanner we have also developed a full 3D face recognition system in the fourth approach (lower right box in Table 1). Note that we did not work on last option (upper right box in Table 1) where the testing images are 3D faces and the training images are 2D.

Table 1. Design space for two-dimensional (2D) and 3D face recognition systems.

		Testing (Verification / Identification)	
		2D	3D
Training (Enrollment)	2D	Most common (Solution 1)	N/A
	3D	Solution 2	Solution 3

The following sections describe each of the face recognition solutions in detail:

3.1 Evidence accumulation for 2D face recognition

Current two-dimensional face recognition approaches can obtain a good performance only under constrained environments. However, in many real-world applications, face appearance changes significantly due to different illumination, pose, and expression. Face recognizers based on different representations of the input face images have different sensitivity to these variations. Therefore, a combination of different face classifiers, which can integrate the complementary information, should lead to improved classification accuracy. We use the sum rule and RBF-based integration strategies to combine three commonly used face classifiers based on PCA, ICA and LDA representations, see Fig. 2. Experiments

conducted on a face database containing 206 subjects (2,060 face images) show that the proposed classifier combination approaches outperform individual classifiers [3].

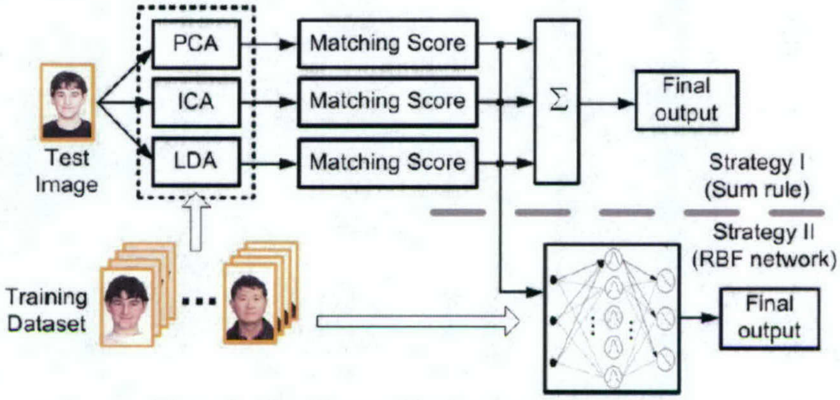


Figure 2. Classifier combination system framework.

A number of applications require robust human face recognition under varying environmental lighting conditions and different facial expressions, which considerably vary the appearance of human face. However, in many face recognition applications, only a small number of training samples for each subject are available; these samples are not able to capture all the facial appearance variations. We utilize a resampling technique to generate several subsets of samples from the original training dataset. A classic appearance-based recognizer, LDA-based classifier, is applied to each of the generated subsets to construct a LDA representation for face recognition. The classification results from each subset are integrated by two strategies: majority voting and the sum rule, see Fig. 3. Experiments conducted on a face database containing 206 subjects (2,060 face images) show that the proposed approach improves the recognition accuracy of the classical LDA-based face classifier by about 7%.

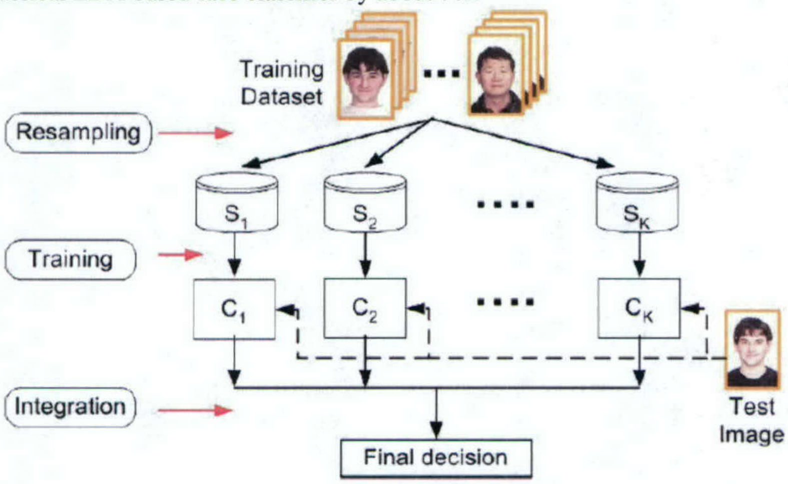


Figure 3. The Resampling-Integration scheme for face recognition. S_1 to S_K are the subsets resampled from the original training dataset. C_1 to C_K are classifiers trained using the corresponding subsets. Here, K is the total number of subsets.

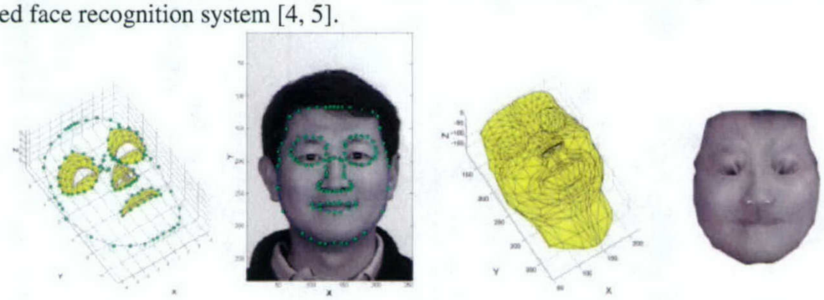
3.2 Demographic information extraction from face images

Human face is a highly rich stimulus that provides diverse information for adaptive social interaction with people. Humans are able to process a face in a variety of ways to categorize it by its identity, along with a number of other demographic characteristics, including ethnicity (or race), gender, and age. Human facial images provide the demographic information, such as ethnicity and gender. Conversely, ethnicity and gender also play an important role in face-related applications. Image-based ethnicity identification

problem is addressed in a machine learning framework. The Linear Discriminant Analysis (LDA) based scheme is presented for the two-class (Asian vs. non-Asian) ethnicity classification task. Multiscale analysis is applied to the input facial images. An ensemble framework, which integrates the LDA analysis for the input face images at different scales, is proposed to further improve the classification performance. The product rule is used as the combination strategy in the ensemble. Experimental results based on a face database containing 263 subjects (2,630 face images, equally split between the two classes) are promising [2], indicating that LDA and the proposed ensemble framework have sufficient discriminative power for the ethnicity classification problem. The proposed scheme can be easily generalized for gender classification. The normalized ethnicity classification scores can be helpful in the facial identity recognition. Useful as a "soft" biometric, the output of ethnicity classification module can be used to update face matching scores. In other words, ethnicity classifier does not have to be perfect to be useful in practice.

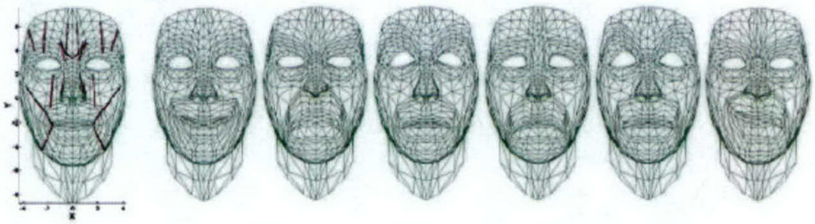
3.3 3D-model enhanced 2D face recognition with a small number of training samples

A robust face recognition system should be able to recognize a face in the presence of facial variations due to different illumination conditions, head poses and facial expressions. However, these variations are not sufficiently captured in the small number of face images usually acquired for each subject to train an appearance-based face recognition system. In the framework of analysis by synthesis, we present a scheme to synthesize these facial variations from a given face image for each subject. A 3D generic face model is aligned onto a given frontal face image. A number of synthetic face images of a subject are then generated by imposing changes in head pose, illumination, and facial expression on the aligned 3D face model. These synthesized images are used to augment the training data set for face recognition. The pooled data set is used to construct an affine subspace for each subject. Training and test images for each subject are represented in the same way in such a subspace. Face recognition is achieved by minimizing the distance between the subspace of a test subject and that of each subject in the database. In our experiments we assume that only a **single** face image of each subject is available for training. Figures 4 and 5 demonstrate the 3D generic model alignment with a 2D intensity image and the synthesis process. Preliminary experimental results show that the proposed scheme is promising for improving the performance of an appearance-based face recognition system [4, 5].



(a) (b) (c) (d)

Figure 4. Face alignment: (a) feature vertices shown as "beads" on the 3D generic face model; (b) overlaid on a given intensity face image; (c) adapted 3D face model; (d) reconstructed images using the model shown in (c) with texture mapping.



(a) (b) (c) (d) (e) (f) (g)

Figure 5. Expression synthesis through 18 muscle contractions. The generic face mesh is: (a) shown in neutral expression (the dark bars represent 18 muscle vectors); distorted with six facial expressions (b) happiness ; (c) anger; (d) fear; (e) sadness; (f) surprise; (g) disgust.

3.4 3D face recognition

In this project we have developed methods for matching 2.5D test scans to a database of 2.5D trainings scans and to full 3D Models. Data for both the models as well as the test scans were captured using a Minolta Vivid 910 3D scanner available in our laboratory. Our results show that using three-dimensional (3D) depth information makes the system more robust to variations in lighting, pose, and facial expression.

We have built two prototype 3D Face recognition systems. The first is written in Matlab and demonstrates the accuracy of our design. The second system is a verification system written in C++. We have achieved three major goals in this project. The first goal is model construction. We designed a method for building a complete model of the surface of our subject's head from a collection of five 2.5D scans. Our second contribution was feature extraction, where we have developed algorithms to automatically find pre-defined anchor points within the scans in order to align the scans with our models. Our third contribution is to build a 3D face matching system that is capable of doing both 3D matching and verification.

Model construction:

The 3D face models are constructed using five 2.5D face scans from different viewpoints. These scans are stitched together using commercially available software, called Geomagic [6] and Rapidform [7]. The models are cleaned up and holes are filled. These models are stored in two formats. VRML is used as a universally transformable format that most 3D modeling software can export. Our own face scan data structure is also used that projects the 3D model on to a cylinder. This projection enables us to write algorithms that can match data much faster.

Using our model construction technique, we have constructed a database of over 100 subject models. The advantage of using a full 3D surface model of the face is that this model is invariant to the pose of the test scan and lighting changes in the environment. Irrespective of the direction the scan was taken, we can still fit it with good accuracy to the complete model.

Feature Extraction:

Our system does not assume that a subject is in a known location looking directly at the camera. Without these assumptions it is difficult to properly align the three dimensional images. Our feature extraction system uses a pose invariant property, called the shape index to help identify possible candidate anchor points. Then a relaxation algorithm searches though the candidate points to find the best set of three anchor points. With the three anchor points the test scan can be properly aligned with the 3D model in a coarse mode.

In order to properly evaluate our algorithms, we have generated a database of over 1,400 face scans from over 100 test subjects [8, 9]. These scans varied in pose direction as well as facial expression and lighting. This highly variable data set helped us push the boundary of face recognition system performance. The results for our feature extractor are quite encouraging. With a database of approximately 600 test scans, we achieved an accuracy of 85.6% when matching a subject's test scan to the same subject's 3D model. To fully understand where the errors are occurring, the test scans were also separated into groups. The following is a list of these groups and the percentage of scans that fall into each group.

Table 2 Test Population Feture Extraction Accuracy Separated by Face Attributes

Attribute	**Population Size (%)**	**Success Rate (%)**
Female	25.2	85.4
Male	74.8	85.7
Facial Hair	11.2	80.6
Dark Skin	10.0	81.7
Eyes Closed	12.0	**98.6**
Asian Features	26.5	84.3
Profile	67.3	79.6
Frontal	32.7	**97.7**
Smile	47.6	82.7
No Smile	52.4	**88.5**

Table 2 shows that facial hair and dark skin make it more difficult to identify key facial features that are needed for alignment. This is a somewhat expected because both of these factors increase the noise produced by the 3D scanner. It is also interesting to note that it is easier to identify feature points in scans with eyes closed than those with the eyes open. This is probably also due to the increase in surface noise that occurs with the eyes open.

Matching:

The recognition engine consists of two components, surface matching and appearance-based matching. The surface matching component is based on a modified Iterative Closest Point (ICP) algorithm. With an initial estimate of the rigid transformation generated from the coarse alignment, the algorithm iteratively refines the transform by alternately choosing corresponding (control) points in the test scan and the 3D model, and finding the best rigid transformation that minimizes an error function based on the distance between them. Our method is a hybrid of two well-known ICP methods [10, 11]. We integrate these two classical ICP algorithms in a zigzag running style. The first algorithm is fast and calculates the distance measure as a point-to-point distance. The second algorithm is more accurate and calculates the point to plane distance. This results in a relatively fast algorithm with a high accuracy. An example of surface matching is provided in Fig. 6.

The candidate list used for appearance matching is dynamically generated based on the output of the surface matching component, which reduces the complexity of the appearance-based matching stage. The 3D model in the gallery is used to synthesize new appearance samples with pose and illumination variations used in the discriminant subspace analysis. The weighted sum rule is applied to combine the two matching components. Experimental results are given for matching a database of 100 3D face models with 598 2.5D independent test scans acquired under different pose and lighting conditions, and some expression changes.

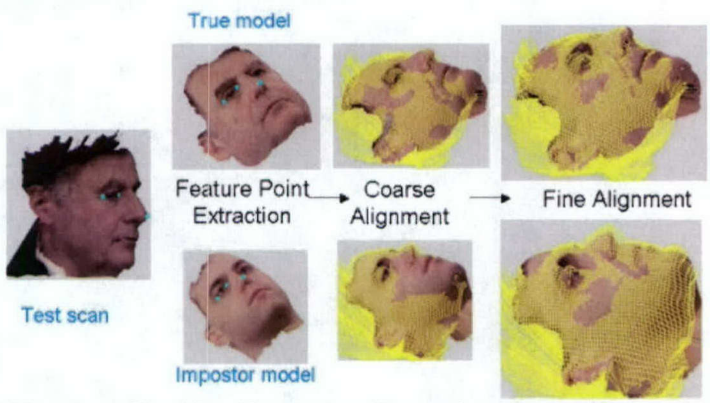


Figure 6. Surface matching streamline. The alignment results are shown by the 3D model overlaid on the wire-frame of the test scan.

The entire face recognition system was tested on 100 3D subject models with a total of 598 test scans. The recognition accuracy is shown in Table 3. A combination of range and intensity data gives better performance than either modality by itself. We are also looking into using deformable face models to better account for changes in expression.

Table 3 Face Recogintion System Accuracy. 3D Face recognition classification accuracy for 100 subject and 598 test scans.

	Classification Accuracy
ICP_Only	87%
ICP + Appearance-based	91%

Table 4 Categorized peformance of rank-one accuracy in recognition.

	Frontal	Profile
w/o smile	99%	98%
w/ smile	78%	85%

Notice that in our test set (see Table 4), a high accuracy is achieved (98% for neutral, 85% for smiling) with the pose variation of approximately 45 degrees from the frontal views. In the recent face recognition vendor test, the reported performance on a data set, where the pose changes are similar to our data set, drops to more than 30% from that of the frontal view matching [12]. This demonstrates the power of 3D models in face recognition applications with large head pose variations.

IV. Summary

This research has made contributions to face detection and recognition. Current approaches to face recognition are mostly based on 2D intensity images. 2D images are not invariant to changes in illumination, facial pose, facial accessories, and expression, resulting in poor face recognition performance. We have developed algorithms that overcome many of these limitations by combining information from different algorithms, utilizing a generic morphable 3D face model and building exact 3D models from a laser scanner.

V. Bibliography

1. Hsu, R.-L., M. Abdel-Mottaleb, and A.K. Jain, *Face Detection in Color Images.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. **24**(5): p. 696-706.
2. Lu, X. and A.K. Jain. *Ethnicity identification from face images. Proc. SPIE.* Orlando, FL, April 2004.
3. Lu, X., Y. Wang, and A.K. Jain. *Combining Classifiers for Face Recognition. Proc. IEEE International Conference on Multimedia & Expo (ICME'03).* Baltimore, MD, July 6-9, 2003.
4. Lu, X., R. Hsu, and A.K. Jain. *Resampling for Face Recognition. Proc. International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA'03).* Guildford, UK June 2003.
5. Lu, X., R. Hsu, and A.K. Jain. *Face Recognition with 3D Model-Based Synthesis. Proc. International Conference on Biometric Authentication, LNCS 3072.* Hong Kong, July 2004.
6. *Geomagic Studio*, http://www.geomagic.com/products/studio/, Raindrop Software.
7. *Rapidform*, http://www.rapidform.com/ 2004, INUS Technology, Inc.
8. Lu, X., D. Colbry, and A.K. Jain. *Three-Dimensional Model Based Face Recognition. Proc. International Conference on Pattern Recognition.* Cambridge, UK, August 2004.
9. Lu, X., D. Colbry, and A.K. Jain. *Matching 2.5D Scans for Face Recognition. Proc. International Conference on Biometric Authentication, LNCS 3072.* Hong Kong, July 2004.
10. Besl, P. and N. McKay, *A Method for Registration of 3-D Shapes.* IEEE Trans. PAMI, 1992. **14**(2): p. 239-256.
11. Chen, Y. and G. Medioni, *Object Modeling by Registration of Multiple Range Images.* Image and Vision Computing, 1992. **10**(3): p. 145-155.
12. *Face Recognition Vendor Test (FRVT)*, http://www.frvt.org/ 2002.