

# Face recognition: evaluation report for FaceIt® Identification and Surveillance

Jingu Heo\*, Besma Abidi, Joonki Paik, and Mongi Abidi.

Imaging, Robotics, and Intelligent Systems Laboratory  
Department of Electrical and Computer Engineering  
The University of Tennessee, Knoxville

## ABSTRACT

The commercial face recognition software FaceIt® Identification and Surveillance was evaluated using the Facial Recognition Technology (FERET) database. The experimental results show the performance of FaceIt® with variations in illumination, expression, age, head size, pose, and the size of the database which all remain difficult problems in face recognition technology.

**Keywords** : FaceIt®, Identification, Surveillance, FERET, illumination, expression, age, pose, size.

## 1. INTRODUCTION

This paper discusses the experimental results from evaluation of the FaceIt® software. FaceIt® is a face recognition software that uses a Local Feature Algorithm (LFA) and is believed to have the highest accuracy of any commercial facial recognition software<sup>1</sup>. We included the results of both the identification, which uses still images, and the surveillance aspects, which use live video input. Two of the most critical requirements in support of producing reliable face-recognition systems are a large database of facial images and a testing procedure to evaluate the system<sup>2</sup>. Therefore, the FERET Database, which includes 14,126 face images from 1,119 individuals with different variations in face images, was used. The FERET Test Procedures are used in evaluation of FaceIt®<sup>1</sup>. We used not only FERET but also our own database (IRIS Lab at UT), which consists of 34 individuals as a gallery and 768 captured faces from 13 individuals as subjects.

A cumulative match characteristic curve (CMC) was used to represent the system's performance; this procedure is a plot of probabilities of correct matches versus the number of best similarity scores. The CMC curve is typically used for identification (one-to-many searches)<sup>1</sup>.

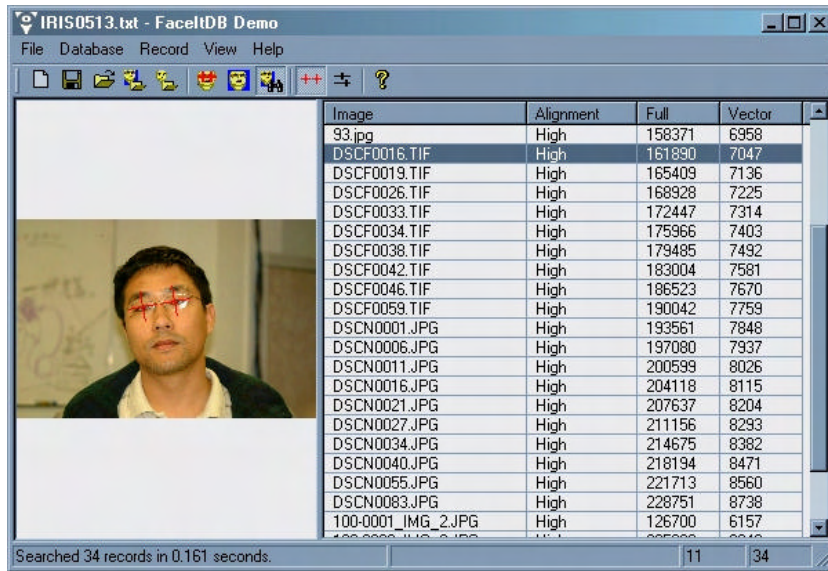
In this paper, we consider the 1<sup>st</sup> match and the within first 10 matches regardless of the database size. Sometimes, it is more reasonable to consider 1% or 5% matching results which reflect the database size.

The remainder of the paper is structured as follows. Section 2 describes the FaceIt® template. Section 3 presents experimental results of FaceIt® Identification considering challenging problems such as expression, illumination, age, pose, and face size. Section 4 shows the experimental results of FaceIt® Surveillance considering lighting conditions, and variations in faces and database sizes. Section 5 concludes by showing directions for future work in face recognition technology.

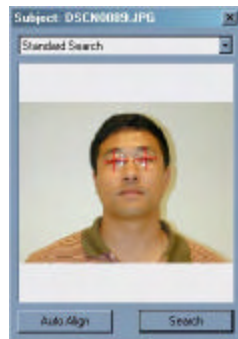
## 2. FACEIT® TEMPLATE WINDOWS

Figure 1 shows the FaceIt® software templates. We can add facial images to the template window in (a) by just dragging the images from the windows explorer. The detailed steps are: First, the gallery database, which includes face images, should be created (a). Then, subject images, which we want to identify, should be chosen, either a still image (b) or a live image(c). Finally, after aligning the subject image and then clicking the search button, the match result can be seen with a confidence rate for each rank (d).

\* E-mail: [jheo@utk.edu](mailto:jheo@utk.edu), Telephone: (865)-974-9685, Fax: (865) 974-5459



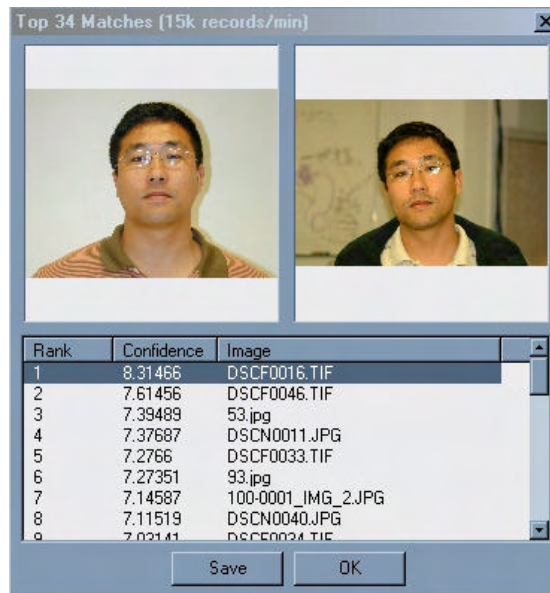
(a)



(b)



(c)



(d)

Figure 1. FaceIt® software templates (a) gallery images or database window, (b) still subject image window, (c) video input window, and (d) matched images sorted by confidence rate.

### 3. FACEIT® IDENTIFICATION EXPERIMENT

In this experiment, we mainly focused on several facial image variations such as expression, illumination, age, pose, and face size. These factors are the main concerns for facial recognition technology. According to the FERET evaluation report <sup>1</sup>, other factors such as compression and media type do not affect the performance and are not included in this experiment.

We divided the evaluation into two main sections with an overall test and a detailed test. In the overall test, we can see the overall accuracy rates of FaceIt®. In the detailed test, we can see what variations affect the system's performance. For lack of databases with mixed variations, we only considered one variation at a time in the facial images for the detailed test. Table 1 shows a summary and description of the tests included in this section. The overall performance of FaceIt® Identification for 1<sup>st</sup> match is about 88%. FaceIt® also works well under expression, illumination and face size variations in cases where these types of variations are not mixed. Age variation is proven to be a challenging problem for FaceIt®.

We did not use all of the images provided by FERET but selected only those suitable for this experiment. The 2-letter codes (fa, fb, and etc) indicate the kind of imagery. For example, fa indicates a regular frontal image. Detailed naming conventions can be seen at [http://www.itl.nist.gov/iad/humanid/feret/feret\\_master.html](http://www.itl.nist.gov/iad/humanid/feret/feret_master.html). Figure 2 and Figure 3 show example images of same individuals under different conditions such as expression, illumination, age, and poses. In the pose test, we still achieve good accuracy rates within  $\pm 25^\circ$  poses. Table 2 and Figure 4 show a summary of the pose tests (R-Right rotation, L-Left rotation). The greater the pose deviations from the frontal view, the less accuracy FaceIt® achieved and the more manual aligning required. Table 3 shows a description of the tests that were not included in this report and the reasons why they were not included. Table 4 shows the execution time and other compatibilities of FaceIt® <sup>3,4</sup>.

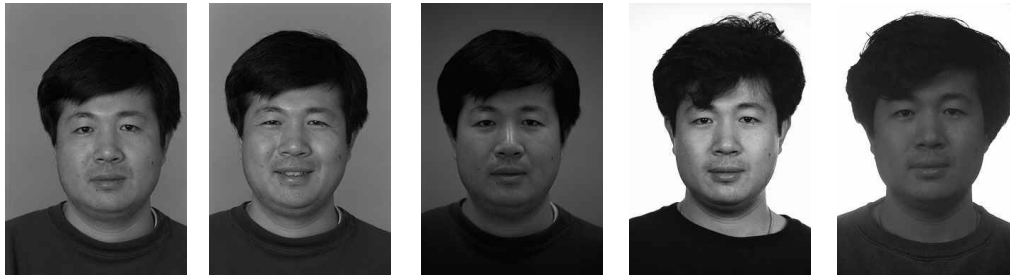


Figure 2. Example Images of same individual under different conditions tested in FaceIt® Identification [Images from FERET].

Table 1. Experimental results for FaceIt® Identification.

Tests	Gallery	Subject	1 <sup>st</sup> Match Success Rate (%)	1 <sup>st</sup> 10 Match Success Rate (%)
Overall Test	700(fa)	1,676(fa, fb)	1,475 (88.0%)	1,577 (94.1%)
Detailed Test				
Expression	200(ba)	200(bj)	197 (98.5%)	200 (100 %)
Illumination	200(ba)	200(bk)	188 (94.0%)	197 (98.5%)
Age	80(fa)	104(fa)	83 (79.8%)	99 (95.2%)
Pose	200(ba)	200(bb ~bh) /pose	Frontal image gives the best result.	
Face Size	200(ba)	200(ba)	No affect as long as the distance between the eyes is more than 20 pixels.	



Figure 3. Example images of same individual under different poses [Images from FERET].

Table 2. Summary of pose test.

Pose(R, L)	First Match (%)	Within 10(%)	Manual Aligning Required (%)
90°L	N/A	N/A	100.0
60°L	34.5	71.0	13.5
40°L	65.0	91.0	4.5
25°L	95.0	99.5	2.5
15°L	97.5	100.0	0.5
0	100.0	100.0	0.0
15°R	99.0	99.5	0.0
25°R	90.5	99.5	2.0
40°R	61.5	87.5	4.5
60°R	27.5	65.0	11.0
90°R	N/A	N/A	100.0

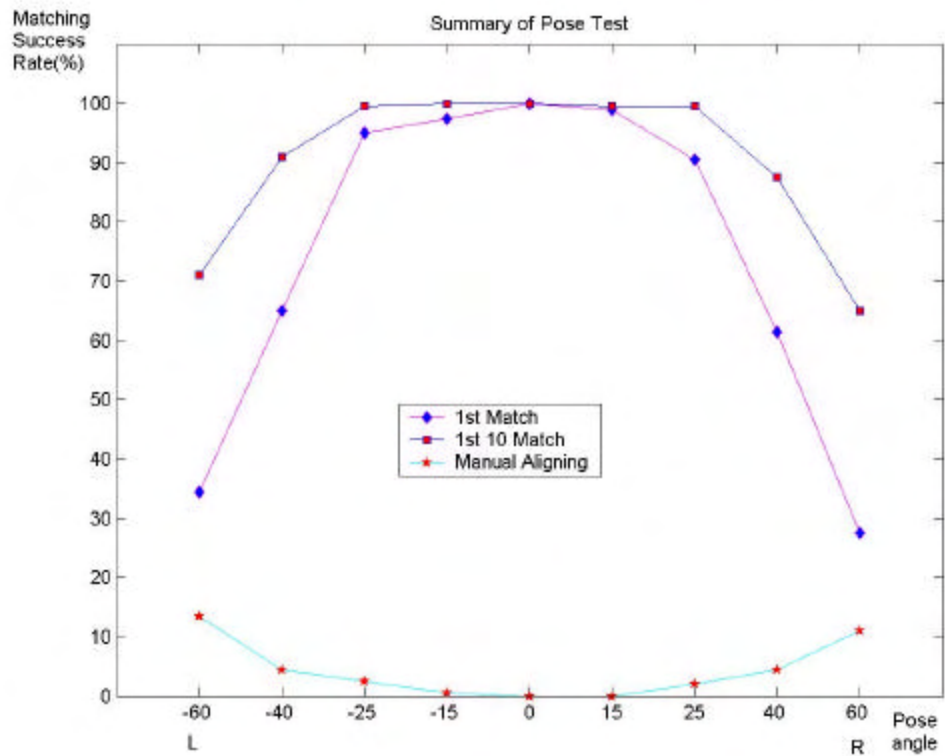


Figure 4. Pose test summary, CMC.

Table 3. Test items not included in this experiment<sup>2</sup>.

Not included	Description	Reason
Compression	Different compression ratios by JPEG	Does not affect performance
Media	Images stored on different media CCD or 35 film	Does not affect performance
Image type	BMP, JPG, TIFF and etc	Does not affect performance
Temporal	Time delay of a photo taken	Covered by overall and age test
Resolution	Image resolution	Features should be seen clearly

Table 4. Execution time and compatibilities.

Feature	Description
Aligning (Eye positioning)	In order to create a gallery database, three steps are necessary; auto aligning, create template and create vector - 2~3 sec / image.
Matching	In order to match against database, subjects should be aligned first (1~2 sec) and then matched (2.5~3 sec; depends on the size of database).
Speed Up	We can load the data onto RAM to speed up process.
Ease of Use	Easy to add and delete images regardless of the size and image types (drag images from the window explorer onto the FaceIt® software).

#### 4. FACEIT® SURVEILLANCE EXPERIMENT

In this experiment, we used a small PC camera from Logitech which was attached to a PC using a USB port to acquire live face images from real scenes. Since we do not have standard test procedures for the surveillance test, we used randomly captured face images and matched these against databases which were used previously in the FaceIt® identification test.

In order to see the effects of variations, we applied different database sizes (the small DB was the IRIS database which contains 34 faces while the large DB was 700 faces from FERET plus the IRIS DB) and lighting conditions to face images. Since face variations are hard to measure, we divided variations such as pose, expressions and age into small and large variations. Figure 5 shows an example of captured faces used in the experiment. When we captured the faces, any person with significant variations such as rotating his head quickly and continuously or notable expression changes was considered as a large variation, while the others were considered as small variations.



Figure 5. Example images of captured faces for FaceIt® Surveillance experiment.

Table 5 provides a result summary for this experiment. The time elapsed between preparation of the IRIS database and the captured faces was approximately 3 months. The basic distance between the camera and the faces was 2~3 ft. The detailed test only used a person who seemed to be moderately well recognized in

the overall test. From detailed tests 1 to 4, we can see how the database size and facial variations affect performance. From detailed tests 3 to 8, we can see how lighting can affect the performance. We can also observe how distance affects the performance from detailed tests 8 and 9. In the lighting conditions, we set 'High' as an indoor ambient illumination condition and 'Medium' as an not ambient but still recognizable through human eyes. 'Front', 'Medium', 'Side', and 'Back' tell the directions of additional lights.

Table 5. Summary of experimental results (basic distance 2~3ft, time elapsed 3 months, Sub: Subject).

Test No.	Description	Gallery DB size	Lighting	Captured Faces /individuals	1 <sup>st</sup> match (Num/Sub)	1 <sup>st</sup> 10 matches (Num/Sub)
Overall 1	Small DB & Large Variations	34	High & Front	758/13	55.8 % (423/758)	96.6 % (732/758)
Detail 1	Small DB & Large Variations	34	High & Front	200/1	55.0 % (110/200)	99.0% (198/200)
Detail 2	Large DB & Large Variations	734	High & Front	200/1	47.5 % (95/200)	78.5 % (157/200)
Detail 3	Small DB & Small Variations	34	High & Front	200/1	67.0 % (134/200)	99.0% (198/200)
Detail 4	Large DB & Small Variations	734	High & Front	200/1	60.5 % (121/200)	93.0 % (186/200)
Detail 5	Small DB & Small Variations	34	Medium	200/1	34.0 % (68/200)	96.5.0% (193/200)
Detail 6	Small DB & Small Variations	34	Medium & Side	200/1	60.5 % (121/200)	98.5% (197/200)
Detail 7	Small DB & Small Variations	34	Medium & Back	200/1	32.0 % (64/200)	80.5 % (161/200)
Detail 8	Small DB & Small Variations Dist: 9~12ft	34	Medium	200/1	0.0 % (0/100)	16.0% (16/100)
Detail 9	Small DB & Small Variations Dist: 9~12ft	34	Medium & Front	200/1	5.0 % (5/100)	78.0 % (78/100)

Figure 6 shows the effects of database size and variations while Figure 7 addresses lighting and distance. A small DB, small variations, close distance, high lighting and additional frontal lighting result in best performance.

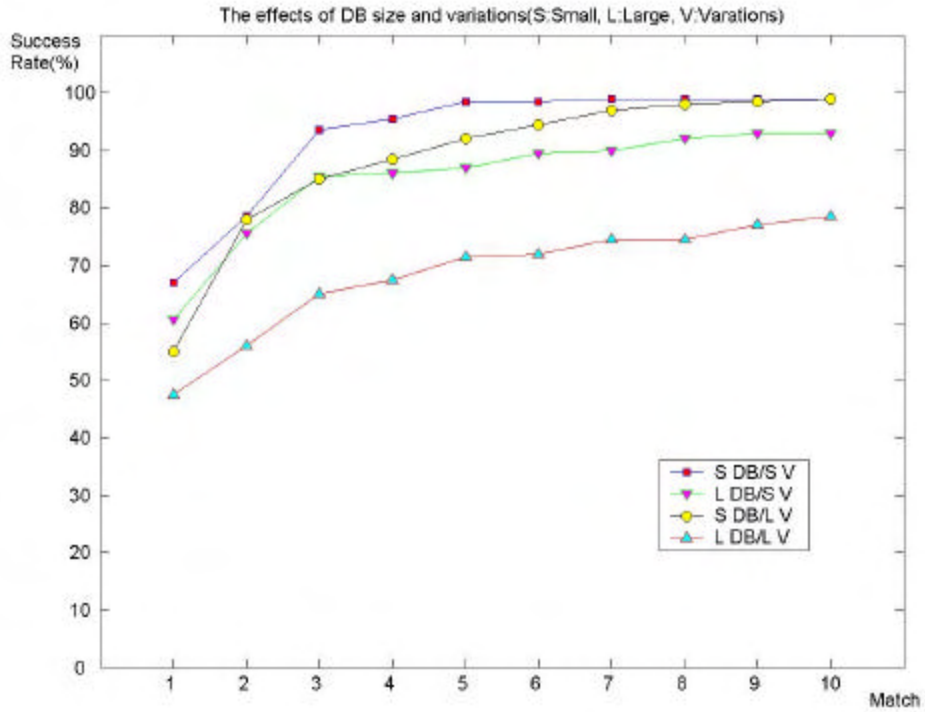


Figure 6. The effects of DB size and variations.

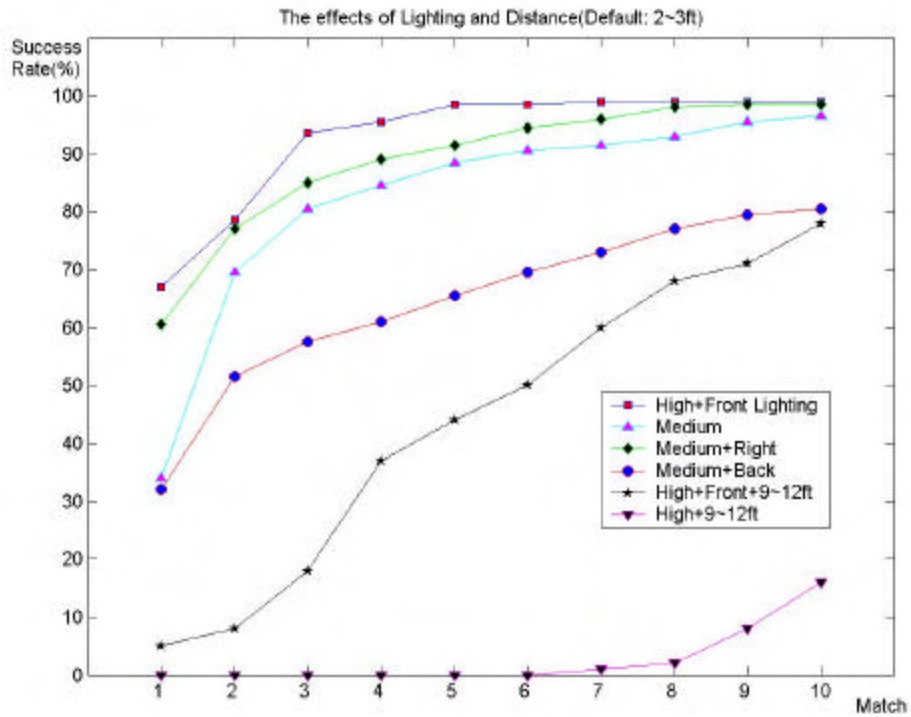


Figure 7. The effects of lighting and distance.

## 5. CONCLUSIONS

We have evaluated FaceIt® and examined how variations in faces affect performance. Local feature based face recognition performance is highly dependent upon the individual. Individuals who have more distinctive features than others can be easily recognized. FaceIt® also better recognizes people with distinct features. This means individuals who have features, which deviate from the average person, can be recognized well and are less affected by variations than an otherwise normal person would be. In real applications where variations occur together, the performance of FaceIt® might be lower than results of our experiment. These variations (especially pose, illumination, and age difference) still present difficult problems in face recognition technology.

Although face-recognition systems work well with “in-lab” databases and ideal conditions, they have been criticized in real applications. So far none of the face recognition systems tested in airports have spotted a single person actually wanted by authorities. They have served only to embarrass innocent people. The technology seems to be better at making innocent incorrect matches, called false positives, than spotting terrorists. Since face recognition is not an exact science but an approach based on user-defined probabilities, the performance can be changed by user-defined thresholds<sup>5</sup> and how to control the area where face images are taken.

In order to increase performance, the methodology for the evaluation of face recognition software should be standardized including a real application test, and include other imaging modalities such as thermal imagery and 3D face modeling which provide more features and features invariant to poses should be developed and incorporated into face recognition systems.

## 6. REFERENCES

1. D.M. Blackburn, J.M. Bone, and P.J. Phillips, “The FERET 2000 Evaluation Report,” *Evaluation Report from NIST*, Feb 2001, pp 1-70.
2. P.J Phillips, H.J. Moon and S.A Rizvi, “The FERET Evaluation Methodology for Face-Recognition Algorithms,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Oct 2000, Vol. 22(10), pp 1090-1104.
3. “FaceIt® Identification SDK manual From Visionics Corporation,” Feb 2000.
4. “FaceIt® Surveillance SDK manual From Visionics Corporation,” Feb 2000.
5. M.Bone and D. Blackburn, “Face Recognition at a Chokepoint: Scenario Evaluation Results,” *Evaluation Report From Department of Defense*, Nov. 2002.