

# 바이오메트릭스 S/W



서연규 / TTA 시험인증연구소 S/W시험인증팀 전임연구원



## 1. 개요

전세계적으로 인터넷이 활성화됨에 따라 사용자에 대한 신뢰성 확보와 보안이 중요하게 인식되면서 보안제품에 대한 수요가 급증하고 있다. 생체정보(바이오메트릭스)를 이용한 보안제품의 경우 분실위험이 없고 쉽게 변형되지 않는다는 점에서 최근에 관심이 고조되고 있다. 생체정보를 이용한 지문인식이나 얼굴인식과 같은 생체인식기술은 범죄수사 분야에 많이 사용되어왔으나, 최근 IT분야와 접목되면서 신분확인 및 통제수단으로써 많은 역할을 할 것으로 보인다. 그 실례로 미국은 테러사건 발생이후 중요 접근통제 수단으로 생체인식기술을 사용하고 있다.

생체인식기술 제품은 지문인식제품이 가장 많은 시장을 차지하고 있지만 홍채, 망막 인식제품도 정확성이 높아 향후 시장관도의 변화가 주목된다. 또한, 음성, 얼굴인식 제품 등도 사용자 편의성 제고를 위해 꾸준히 연구되어 왔으며 다양한 분야와 접목되고 있다. 각 제품이 가격, 편리성, 응용분야 등 특성을 가지고 있지만 생체인식기술 제품의 시장확대 및 범용화를 위해서는 신뢰확보가 필수적이다. 또한, 개인 생체정보를 이용하는 것이므로 접촉에 대한 거부감을 유발할 수 있는데 이를 해소하는 방법도 모색되어야 할 것이다. 신뢰성이란 제품에 대한 사용에 있어 정확하고 안전한 방법을 제공하는 수단을 의미한다. 이러한 정확성 및 안전성을 확보하기 위해서는 공정하고 객관적인 평가방법이 보장되어야 한다.

본 고에서는 생체정보를 이용한 분야로서 생체인식기술 제품의 시험인증기술 동향을 파악하고자 한다. 이를 위해 국내외 기관에서 수행한 생체인식기술 제품평가프로젝트들에서 사용한 평가항목, 평가데이터 수집방법 및 평가방법 등을 살펴본다.

## 2. 국내외 시험인증기술 동향<sup>1)</sup>

생체인식기술 제품에 대한 평가방법은 크게 성능평가와 보안성 평가로 구분할 수 있는데, 성능평가는 사용자에게 편리성 및 정확성을 보장하기 위한 것이며, 보안성 평가는 제품에 대한 안전한 사용을 지원하는 것이다. 성능평가의 경우, U.K Biometric Working Group(BWG)에서 제시하고 있는 'Best Practices'가 일반적인 시험방식으로 사용되고 있으며, 보안성 평가의 경우 영국(U.K, BWG) 및 미국(DoC, Biometric Management Office)에서 Common Criteria (CC)에 따른 Biometric 'Protection Profile'에 대한 기준을 제안하고 있다. 국내에서는 한국정보보호진흥원(KISA)에서 2001년부터 2003년까지 "생체인증시스템 보안성 평가기술 개발" 프로젝트를 통하여 CC기반의 지문인식제품 보안성 세부평가기준을 개발하였다. 한국정보통신기술협회(TTA)에서도 2004년부터 생체인식기술 제품 시험인증을 위한 평가모델 개발을 수행하고 있다.

생체인식기술 제품을 평가하기 위해서는 시스템 특성에 근거한 평가항목을 도출하고 평가항목에 따라 시스템을 평가할 수 있도록 평가데이터를 수집한다. 특히, 생체인식기술 제품의 경우 수집된 데이터가 평가결과에 미치는 영향이 크므로 평가데이터 수집방법은 매우 중요하다. 본 절에서는 생체인식기술 제품의 평가프로젝트인 Sandia National Laboratory, INSPASS(Immigration and Naturalization Service Passenger Accelerated Service System), SSS(Social Security System), FVC2000((Fingerprint Verification Competition 2000), FVC2002, FRVT(Face Recognition Vendor Test), FERET(FacE REcognition Technology), BIOTEST, BIOIS(BIometric Identification System) 등에서 공개된 평가항목, 평가데이터 수집방법, 평가방법에 대해 살펴본다.

### 1) 평가항목

바이오메트릭스 시스템 평가를 위해 주로 사용되는 항목은 [표 1]과 같다

[표 1] 프로젝트에서 사용된 평가항목

평가항목	평가 내용
FRR(False Rejection Rate) 또는 FNMR(False Non Match Rate)	데이터가 일치하나 거부하는 비율
FAR(False Acceptance Rate) 또는 FMR(False Match Rate)	데이터가 일치하지 않으나 수락하는 비율
평균 등록시간	데이터를 등록하는데 소요되는 평균 시간
평균 정합시간	데이터를 정합하는데 소요되는 평균시간

1) 2003년 "생체인증시스템 보안성 평가기술 개발" 연구결과와 관련 참고문헌 내용을 일부 발췌함

평가항목	평가 내용
ROC Curve	FAR에 따른 FRR의 변화비교
등록 거부율(Failure to Enroll)	데이터입력 후 템플시 생성이 실패하여 등록거부된 비율
시스템 처리능력(System Throughput)	대용량 데이터베이스 검색에 따른 부하처리 능력

이외에도 FVC2000, FVC2002에서는 EER(Equal Error Rate : 매칭 실패가 제외된 FMR 과 FNMR에 의해 계산), ZeroFMR(FNMR이 0인 경우 중에서 가장 낮은 FNMR 값), ZeroFNMR(FMR이 0인 경우 중에서 가장 낮은 FMR 값)을 사용하였으며, SSS에서는 Penetration rate(검색효율), Binning error(오분류) 등이 사용되었다. BIOIS 프로젝트에서는 실질적인 평가대상은 아니지만 Setup time, Acquisition Time, Recognition Time 등의 운영에 필요한 시간도 고려하였다.

얼굴인식의 경우 FRVT(Face Recognition Vendor Test)에서 인식성능 평가(Recognition Performance Test)를 위해 [표 2]와 같은 평가항목을 사용하였다.

[표 2] FRVT의 인식성능 평가를 위한 평가항목

평가항목	평가내용
Compression Experiments	영상 압축시 압축 손실에 대한 인식을 테스트
Distance Experiments	모델과 카메라와의 거리에 따른 인식을 테스트
Expression Experiments	표정변화에 따른 인식을 테스트
Illumination Experiments	조명변화에 따른 인식을 테스트
Media Experiments	비교가 되는 영상이 서로 다른 영상매체를 이용하여 획득되었을 경우 인식을 테스트
Pose Experiments	포즈변화에 따른 인식을 테스트
Resolution Experiments	영상의 사이즈 변화에 따른 인식을 테스트
Temporal Experiments	시간 변화에 따른 인식을 테스트

## 2) 평가데이터 수집

평가를 위해서는 데이터 수집은 평가결과에 신뢰성을 주는 가장 중요한 요인이다. 지문의 경우는 습도 및 온도를 고려하여야 하며, 얼굴의 경우 포즈, 표정, 조명등의 변화에 영향을 받지 않도록 다양하고 대표성 있는 데이터 수집이 중요하다.

INSPASS에서는 손모양 인식제품 평가를 위해, 2946 templates, 9862 transactions을 사용하였고, template 생성을 위해 3회 반복을 통하여 오른손 데이터를 획득하였다. 각 데이터는 0~255의 범위를 갖는 9 integers에 대한 9-byte vector로 표현하였고, vector 값은 single 14 character glyph로 encoding을 통해 문자크기를 줄였다.

SSS에서는 지문인식 평가데이터 수집을 위해 남녀비율을 각각 5.5:4.5로 하고 Identicator DF-90 “flat” scanner를 사용하여 1인당 8개의 엄지와 검지지문을 등록하여 TIFF로 압축하였다. 데이터 수집시 training, practice, test image의 집합으로 분류하여 구성하였다. training data의 경우, 습도를 반영하여 510명의 지원자로부터 4080개의 이미지를 획득한 후 가장 좋은 이미지만으로 데이터 집합을 생성하였다. test data는 training data에 참여한 409명을 포함한 506명으로부터 4128개의 이미지를 획득하였으나 이미지 품질에 대해서 다소 방관하였다. 또한 10명의 지원자로부터 시간적 간격을 두고 10회의 중복된 이미지를 획득하였다. practice data는 training data에 참여한 10명으로부터 80개의 이미지 집합을 획득하여 사용하였다.

FVC2000에서는 지문인식데이터를 생성하기 위해 4가지의 서로 다른 종류의 DB를 구축하였는데 [표 3]은 이들 데이터베이스의 특징을 보여주고 있다.

[표 3] FVC2000에서 사용한 4가지 DB의 특징

	센서 타입	이미지크기	Set A(w × d)	Set B(w × d)	해상도
DB1	Low-cost Optical Sensor	300 × 300	100 × 8	10 × 8	500 dpi
DB2	Low-cost Capacitive Sensor	256 × 364	100 × 8	10 × 8	500 dpi
DB3	Optical Sensor	448 × 478	100 × 8	10 × 8	500 dpi
DB4	Synthetic Generator	240 × 320	100 × 8	10 × 8	약 500 dpi

BWG의 ‘Best Practice’에서는 합성이미지 사용방법에 대해서 고려하지 않는다. 합성이미지는 알고리즘이 처음 접하는 이미지를 제공함으로써 알고리즘간의 고정 비교가 가능하고, 매우 낮은 비용으로 이미지를 비교적 빠르게 생성할 수 있다. 또한, 합성이미지는 개인정보 유출 등의 개인 프라이버시 문제를 해결할 수 있으며, 최대 회전, 변환, 피부왜곡의 양 등을 조절함으로써 평가 데이터베이스의 난이도 조절이 가능하다는 장점이 있다. 합성이미지가 실제 지문이미지를 재현하는데 적합한지의 여부는 3개의 실제 지문이미지 데이터베이스를 통해 검증이 가능하다. 합성이미지 생성시 생성기의 파라미터는 작은 지문 채취 영역을 갖는 low-cost 센서를 모방하기 위해 튜닝하였다. 앞의 3개의 데이터베이스를 사용하여 최대 회전과 이동, 그리고 피부 왜곡 등을 조절하여 혼동을 줄 수 있는 지문을 대략적으로 생성하였다.

얼굴인식 제품을 평가하기 위한 FERET(FacE REcognition Technology) 프로젝트는 FERET 데이터베이스(FERET DB)를 통해 대량의 데이터베이스에 대한 알고리즘의 성능파악에 중점을 두고 있다. 따라서 데이터 수집시 연령과 인종, 성별에 대한 비율은 정하지 않았으며 정면, 좌측, 우측 등의 포즈변화와 표정, 조명 변화 등을 고려하여 컬러 영상을 획득하였다. 획득된 이미지는 8 bit 흑백 영상으로 변환한 후 TIFF 파일 포맷으로 저장하였다. 이미지 크기는 256 × 384로 하였고, 카메라 위치를 조정하여 영상 내 얼굴 크기의 가변성을 두었으며, 머리와 목 그리고 가깝게 어깨 등이 영상 내에 포함되어 있다. 영상 획득 시 추가적인 환경 요소로는 안경과 헤어스타일, 조명 등이 있는데, 안경과 헤어스타일을 제외하고는 영상 획득 후, MATLAB Image Processing Tool Box를 이용하여 일부영상에 대해 조명을 40%와 60%를 각각 줄였다. 그리고

스케일(scale)도 일부영상에 대해 10%, 20%, 30%를 각각 줄였으며, Adobe Photoshop을 이용하여 모델이 착용한 옷의 색깔을 바꾸었다. FRVT2000에서는 FERET 데이터베이스와 NIST의 Human ID 데이터베이스 외에 Indoor Still Image Station, Outdoor Still Image Station, Indoor and Outdoor Video Image Station 등을 추가로 사용하였다.

### 3) 평가방법

평가는 전문적인 평가항목들을 객관적으로 측정하기 위한 방법으로 주요 평가항목인 FRR과 FAR 측정을 위해 각 사용자에 대한 genuine attempt와 imposter attempt를 각각 여러 차례 반복 수행하였다. 또한 genuine과 inter-template에 대한 distance distribution을 이용하거나 imposter 계산을 위한 non-matching template을 랜덤하게 할당할 수 있는 방법이 없는 경우 inter-template distribution을 이용해서 근사치를 계산하는 등 수학적 방법도 사용하였다.

FERET에서 사용된 평가방법은 Phase I, Phase II, Sep96 evaluation로 구분된다. Phase I은 대용량의 데이터베이스에 대한 알고리즘 성능을 측정하기 위한 ‘대용량 갤러리 테스트’, 갤러리 수가 프루브(Probe)에 비해 절대적으로 적을 때 성능을 측정하기 위한 ‘False-Alarm 테스트’, 갤러리와 프루브에 있는 각각의 영상이 다른 포즈를 하고 있을 때의 성능을 측정하기 위한 ‘회전테스트’ 등으로 구성된다. Phase II는 Phase I과 동일하나 사용된 DB가 갤러리 세트에서 831장으로 구성되어 있고, 463장의 중복된 영상이 프루브 세트에 있다는 것이다. 그리고 테스트는 대용량 갤러리 테스트에 대해서만 수행되어졌다. 마지막으로 Sep96 evaluation은 세 가지 평가 원리를 정하고 있는데, 첫 번째, 동일인으로부터 획득된 영상들을 타겟세트와 쿼리세트에 분배를 하되 동일영상이 타겟세트와 쿼리세트에 존재하지 않도록 하여 타겟세트의 영상은 DB 내에서 유일한 영상이어야만 한다. 두 번째, 학습(등록) 시기를 테스트 이전에 두어 학습이 먼저 완료 되어야만 한다. 세 번째, 모든 알고리즘들은 두 장의 영상(타겟세트와 쿼리세트에서 선택) 사이에 대해서만 유사도를 측정한다는 것이다.

FRVT2000에서는 IEEE 논문 “An Introduction to Evaluating Biometric Systems”<sup>2)</sup>에 기반하여 평가를 수행하였다. 이 방법은 크게 Technology Evaluation, Scenario Evaluation, Operational Evaluation 등 세 단계로 구분되며 다시 Recognition Performance Test와 Product Usability Test 두 가지로 구분하였다. Recognition Performance Test는 1)에서 전문적인 평가항목을 테스트 하는 것이며, Product Usability Test는 업체에서 제출한 기존 템플릿과 새로운 템플릿을 이용하여 사용자의 카메라 응시여부에 따른 Verification Test, Identification Test, 조명변화에 따른 Backlight Test로 구성된다.

2) P.J. Phillips, A. Martin, C.L. Wilson, and M. Przybocki, “An Introduction to Evaluating Biometric Systems”, 2000, IEEE Computer Society Press, pp. 56~63

### 3. 결론

지금까지 생체정보를 이용한 생체인식기술 제품의 평가를 위해 국내외 프로젝트들에서 사용한 평가항목, 평가를 위한 데이터 수집방법 그리고 평가방법 등을 살펴보았다. 평가항목은 주로 FAR, FRR, 정합시간 등 다양한 평가항목이 사용됨을 알 수 있었고, 이러한 평가항목을 평가하기 위해 데이터셋의 구성과 평가방법을 보았다. 또한 바이오메트릭스 시스템 평가에 중요한 사항으로 데이터 수집은 남녀비율, 습도, 온도, 표정 및 포즈 등, 다양성과 대표성을 고려하였음을 알 수 있다.

이러한 평가기술은 생체인식 기술의 신뢰성을 높이고 시장 보호와 확대에 중요한 역할을 한다. 이를 위해서는 공인된 평가기관과 정확하고 공정한 평가기술 확보가 필요하다. 세부 알고리즘의 성능 및 보안성을 위한 정량적 평가기준과 방법이 필요하다. 또한 업체에서 평가결과를 객관적으로 받아들일 수 있도록 생체특성의 변화와 입력 또는 인증방식 등의 환경변화에 따라 최적의 평가환경을 제공하여야 한다. 더 나아가서는 생체특성별 평가에 사용할 수 있는 표준 데이터 베이스를 구축하고 관리하여야 할 것이다. **TTA**