

얼굴인식시스템 성능평가 도구의 설계 및 구현

신 우 창*

The Design and Implementation of a Performance Evaluation Tool for the Face Recognition System

Woo Chang Shin*

■ Abstract ■

Face recognition technology has lately attracted considerable attention because of its non-intrusiveness, usability and applicability. Related companies insist that their commercial products show the recognition rates more than 95% according to their self-testing. But, the rates cannot be admitted as official recognition rates. So, performance evaluation methods and tools are necessary to objectively measure the accuracy and performance of face recognition systems. In this paper, I propose a reference model for biometrics recognition evaluation tools, and implement an evaluation tool for the face recognition system based on the proposed reference model.

Keyword : Face Recognition, Biometrics, Performance Evaluation, Reference Model

1. 서 론

생체인식 시스템이란 측정할 수 있는 개인의 신체적 특성 또는 습관적인 행위를 이용하여 기록된 내용과 비교함으로써 개인의 신분을 확인하거나, 특정한 개인을 여러 등록된 특성 정보 중 찾아내는 장비 또는 그러한 과정을 의미한다. 생

체인식 시스템에는 지문인식, 홍채인식, 정맥인식, 얼굴인식 등이 있으며, 이 중 얼굴인식은 특별히 얼굴의 특성을 이용하는 것을 말한다.

이러한 얼굴인식시스템은 다른 생체 인식시스템과는 다르게 사용자 신체의 일부를 인식 장치에 의도적으로 접근하지 않아도 되어 생체 정보의 획득에의 거부감과 불편함이 적을 뿐만 아니라 기존

* 서경대학교 인터넷정보학과

에 널리 사용되는 영상 인식 장비를 그대로 사용할 수 있다는 점에서 비교적 쉽고 저렴하게 대량의 생체정보를 획득할 수 있는 장점이 있다. 특히 미국의 9.11 테러사건 이후 각국의 공항이나 항만과 같은 입·출국 관련 시설에서 얼굴인식시스템의 도입과 같이 그 시장성이 커짐에 따라 다양한 얼굴인식 알고리즘과 상용 시스템이 개발되어 제안되고 있다. 이와 같이 다양한 이점과 필요성으로 인해 얼굴인식시스템에 관한 사회적 요구가 증가함과 더불어, 이러한 얼굴인식시스템을 평가할 수 있는 공인된 성능평가가 중요해 지고 있다.

외국에서는 얼굴인식 기술의 발전을 위하여 FERET(FacE REcognition Technology) 이나 FRVT(Face Recognition Vendor Test) 같은 체계적인 성능평가 프로그램들을 마련하고 실행하고 있는 반면, 국내의 얼굴인식 기술의 수준은 아직 초기단계로 국제경쟁력을 확보하지 못하고 있다. 일반적으로 개발된 얼굴인식시스템의 경우 대부분 자체적인 테스트를 거쳐 95% 이상의 인식률을 보고하고 있으나, 이는 각 시스템에 유리한 가정과 특정한 얼굴 데이터베이스를 사용하였기 때문에 객관적으로 신뢰하기 어렵다. 따라서 얼굴인식시스템들 간의 정확성과 성능을 객관적으로 비교·평가할 수 있는 표준화된 얼굴데이터베이스와 성능평가 방법 그리고 성능평가 도구가 필요하다[2].

성능평가 도구를 활용하여 성능평가 항목과 절차에 따라 얼굴인식시스템의 성능을 평가함으로써, 1) 개발자 입장에서는 개발된 기술에 대한 객관적 평가를 얻을 수 있어서, 기술 개발의 지표로 활용할 수 있으며, 2) 얼굴인식시스템 구매자 입장에서는 자신이 필요한 조건에 적합한 기술과 제품을 구매하는데 도움을 받을 수 있다. 또한 3) 객관적인 성능시험과 공신력 있는 인증을 통하여 얼굴인식시스템 시장의 활성화에 기여할 수 있다.

본 논문에서는 생체정보 인식시스템 성능평가를 위한 참조모델을 제안하고, 제안된 참조모델을 바탕으로 얼굴인식시스템의 성능을 비교·평가할

수 있는 성능평가 도구를 설계 및 구현한다. 참조모델은 관련 국제 표준안들과 호환이 되도록 설계되었으며, 이를 바탕으로 개발된 성능평가 도구의 일관성과 신뢰성 향상에 기여한다.

제 2절에서는 얼굴인식시스템 성능평가에 관련된 기존 연구들에 대하여 살펴보고 제 3절에서는 얼굴인식시스템 성능평가를 위한 참조모델을 제안한다. 제 4절에서, 제안된 참조모델을 바탕으로 성능평가 도구의 설계 및 구현에 대하여 기술한다. 제 5절에서 실제 전문업체의 얼굴인식 모듈을 이용하여 성능평가를 수행한 예를 보이며, 제 6절에서는 본 논문에서 제안한 성능평가 도구를 이용한 성능평가 방법과 기존 성능평가 프로그램들을 비교한다. 마지막 제 7절에서 결론 및 향후 과제에 대하여 살펴보고 마무리를 짓는다.

2. 관련 연구

얼굴인식시스템에 관한 대표적인 공인 성능평가 프로그램으로 FERET(FacE Recognition Technology)과 FRVT(Face Recognition Vendor Testing)가 있다. 1993년부터 시작된 FERET은 얼굴인식시스템의 성능평가뿐 아니라 알고리즘의 개발 및 얼굴인식용 데이터베이스의 수집까지 포함하는 것이었다. FERET은 미국 국방성 주도로 1993년부터 단계적으로 1997년까지 수행된, 조명, 크기, 자세, 배경 등의 환경 변화와 사진 획득 시간의 차이, 그리고 대용량 데이터베이스 처리에서의 알고리즘의 성능을 시험한 평가이다. 특히 FERET의 성능평가는 연구소 수준의 알고리즘을 측정하기 위해 고안된 일반적인 평가였고, FERET 성능평가의 주요 목표는 얼굴인식의 최신 기술과 유연성에 적응하는 것이었다. 따라서 FERET 테스트는 알고리즘의 개별 컴포넌트의 성능에 대한 영향을 명백하게 측정하기 위한 것도 아니었고, 시스템이 동작하는 모든 상태에서 잘 정렬된 과학적 방법으로 성능을 측정하기 위한 것도 아니었다 [9-11, 16].

FRVT는 세 번의 FERET 성능평가(1994년, 1995년, 1996년)를 기반으로 수행된 얼굴인식시스템 성능시험 평가이다. FERET 프로그램이 얼굴인식 분야에 평가기법을 도입하고 시스템의 프로토타입 단계인 초보적 수준에서 얼굴인식 분야를 발전시킨 역할을 한 반면, 2000년에 얼굴인식 기술이 프로토타입 수준에서 상용 시스템 수준으로 성숙함에 따라, 2000년도에 FRVT 2000이 이러한 상업 시스템의 성능을 측정하고 마지막 FERET 평가와 비교하여 기술적 진보가 얼마나 되었는지를 평가하였다. 따라서 FRVT는 상업적인 이용 가능성에 대한 기술적인 평가를 했다는 데 그 의미가 있다고 할 수 있다.

또한 2002년에 와서는 얼굴인식 기술에 대한 일반인의 관심이 대폭 증대되었다. 따라서 FRVT 2002의 경우 2000년 이후의 기술적인 발전 정도를 측정하고, 실제로 쓰이는 큰 규모의 데이터베이스에 대해 평가하고, 얼굴인식 성능을 더 잘 이해하기 위한 새로운 실험을 도입하였다. 이러한 일련의 성능평가는 평가 이론뿐 아니라 얼굴인식 기술이 성장함에 따라 평가의 규모나 난이도 그리고 복잡성이 점점 증가해 왔다. 예를 들면 FERET Sep96의 경우 72시간 동안 1,450만 번의 비교를 했고, FRVT 2000은 72시간 동안 19,200만 번의 비교를 했던 반면 FRVT 2002는 264시간 동안 150억 번의 비교를 하는 평가를 도입했다. FRVT 2002는 또한 이전보다 더 세분화 된 항목으로 갤러리(gallery)를 나누어 실험을 하였다. 이 실험은 인증(verification), 식별(identification), 그리고 감시목록(watch list)의 얼굴인식 성능평가의 기본 세 가지 작업에 대하여, 표정이나 조명, 시간의 변화, 실내/실외, 포즈 등의 기본적인 실험 항목들과, 얼굴의 해상도, 이미지 압축 방법, 미디어의 종류, 카메라에서 얼굴까지의 거리, 갤러리의 크기, 순위(rank)함수와 같은 상세 분석 항목, 그리고 3차원 모퍼블(morphable) 모델이나 평준화(normalization), 비디오 같은 기술적 평가 항목 그리고 통계적 요소 항목인 성별, 나이에 따른 항

목에 대해 평가하였다[5, 8, 13, 14].

FERET과 FRVT와 같은 공인 성능평가 프로그램은 주로 얼굴인식시스템의 알고리즘 정확도를 측정하기 위한 프로젝트였으며, 시험을 위한 얼굴영상 데이터베이스가 공통으로 제공되고 각자의 방법으로 일정 기간 동안 얼굴인식을 수행하고 이 결과를 가지고 평가하는 방식으로 진행되었다. 그러나 이러한 방식은 우선 해당 기간 동안 참여한 얼굴인식시스템 업체에 한정되어 평가 받을 수 밖에 없고, 또한 특정 얼굴영상 데이터베이스를 사용함으로써 고정된 데이터베이스에 의한 평가가 될 수 밖에 없다. 또한 데이터베이스의 항목들이 영상의 크기, 대상의 자세, 영상 획득 환경, 시간 등에 한정되어 있어 알고리즘을 평가하는 다양한 조건들을 유동적으로 만족시키지 못하는 문제가 있다. 그리고 이러한 알고리즘을 평가하기 위한 환경을 각 얼굴인식시스템 개발 업체들에 맡김으로써, 성능평가 시스템 환경 구축이 통일되지 않는 문제도 존재한다. 또한 알고리즘의 수행 결과를 다시 분석하고 각 알고리즘의 정확도를 판단하는 문제에 있어서도 부가적인 작업이 필요한 방식으로 진행된 것이 사실이다. 따라서 이러한 문제들을 해결할 수 있는 알고리즘 평가 방식의 설계와 표준화된 평가 환경의 구축, 그리고 이러한 환경을 바탕으로 성능 측정된 알고리즘의 평가 결과를 자동으로 도출하고 이를 바탕으로 알고리즘의 인증 여부도 결정할 수 있는 방식을 사용하여야 할 필요성이 제기된다.

국내에서는 얼굴인식시스템에 대한 공식적인 성능평가 프로그램이 실행된 적은 아직 없으나, 관련 연구는 꾸준히 수행되어 왔다. 한국정보보호진흥원에서 2002년도에 얼굴인식시스템 성능평가에 관련된 외국사례를 분석하고 평가방법론을 제시한 “얼굴인식 시스템 성능평가 방법론 연구”[3]가 수행되었으며, 2004년도에 성능평가 방법과 성능시험도구를 개발하는 “얼굴인식시스템 성능시험도구 개발” 과제를 수행하였다[2]. 이들 연구과제는 얼굴인식시스템의 성능시험을 위한 방법을

제시하고 이를 국내 처음으로 구현함으로써 객관적인 성능시험의 모습을 보여주었으나, 실제 외부 업체의 얼굴인식시스템을 대상으로 성능평가를 하지 못하고, 과제 내에서 개발된 자체 얼굴인식 모듈에 대해서 성능평가를 실행함으로써 객관적인 성능평가 도구로서의 면모를 보여주지 못하였다. 본 논문에서 제안하는 성능평가 도구는 한국 정보보호진흥원에서 2006년도에 수행된 “얼굴인식시스템 성능시험도구 기능개선 연구” 과제의 성과물이다.

3. 참조 모델

공정한 생체정보 인식시스템 성능평가 시스템을 구축하기 위해서는 고려해야 될 사항이 많다. 얼굴인식시스템 성능평가를 예로 들면 얼굴인식에 이용될 얼굴데이터베이스 수집과 함께 성능평가 테스트항목-표정변화, 조명변화 등-을 정해야 하고, FAR, FRR, EER 과 같은 성능평가 측정항목 역시 결정하여야 한다. 또한 평가대상이 되는 얼굴인식시스템과 성능평가 시스템과의 표준화된 인터페이스를 설계해야 하며, 공정성과 신뢰

성을 높이기 위하여 성능평가 각 단계에서 국제 표준을 준용할 필요가 있다.

이렇듯 성능평가 시스템 구축에 있어서 고려해야 될 사항들을 분석하고 정리하여 성능평가 시스템 개발에 도움을 주기 위해 만든 것이 참조모델이다. 참조모델은 성능평가 시스템을 구축하기 위한 시스템 기본구조와 가이드라인, 그리고 성능평가 시스템 구축 프로세스를 제시한다.

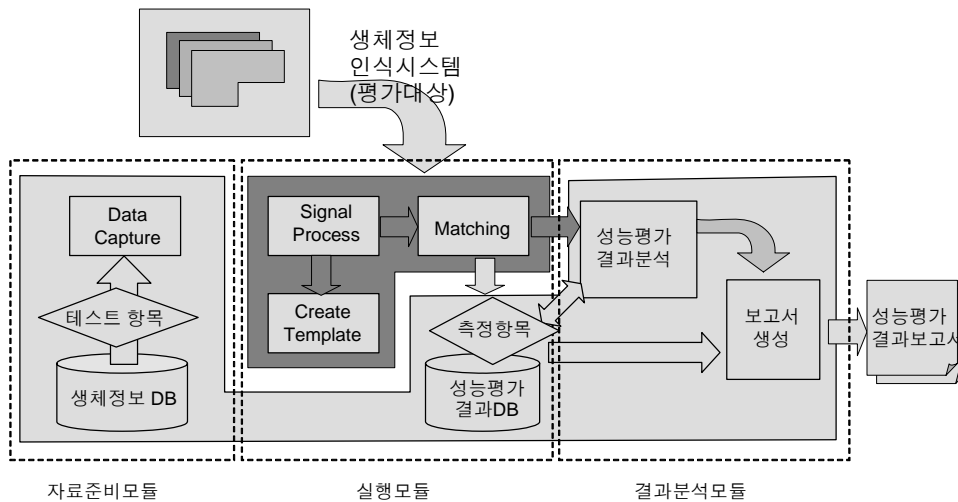
본 논문에서 제안한 참조모델은 1) 생체정보인식기술의 알고리즘 성능평가를 위한 것이며, 또한 2) 생체정보 인식시스템과 연동하여 자동적으로 성능평가를 수행하고 결과를 출력하는 시스템을 구축하기 위한 것이다.

3.1 참조모델 구조

생체정보 인식기술 알고리즘 평가시스템을 위한 참조모델 구조는 자료준비모듈과 실행모듈 그리고, 결과분석모듈로 구성된다. 참조모델 구조는 [그림 1]과 같다.

3.1.1 자료준비모듈

자료준비모듈은 성능평가에 이용될 생체정보



[그림 1] 참조모델 구조

(영상, 동영상, 음성 등)를 준비하는 부분으로, 생체정보데이터베이스 구축과 테스트항목 설계가 주요 고려대상이 된다.

생체정보 데이터베이스 구축은 생체정보인식 분야의 연구에 핵심적으로 필요한 부분으로, 외국의 경우 다수의 FERET이나 FRVT 프로그램을 진행하면서 상당한 양의 생체정보 데이터베이스를 체계적으로 구축해왔다. 국내에서도 한국정보보호진흥원에서 관련 연구를 수행하여, 한국형 생체정보 데이터베이스를 구축하였다[4]. 어떤 생체정보 데이터베이스를 이용하여 성능평가를 수행하는가에 따라 성능평가 시스템의 평가 신뢰성에 큰 영향을 미치게 되기 때문에, 시스템 개발 초기에 우선적으로 고려해야 될 사항이다. 또한 평가의 신뢰성을 위해서는 성능평가에 사용되는 생체정보는 외부에 공개된 적이 없는 새로운 데이터이어야 한다[12].

일반적으로 생체정보인식기술의 알고리즘 성능평가는 표준적인 생체정보인 갤러리를 학습 또는 등록시킨 후 인식대상이 되는 테스트 생체정보인 프로브(probe)와 비교하여 둘 간의 유사도를 측

정한다. 이때 둘 간의 정보생성 환경이나 조건에 따라 알고리즘 성능이 달라지게 된다. 예를 들면, 얼굴인식의 경우 갤러리로 무표정한 정면 얼굴영상을 등록한 후, 프로브로 웃는 표정의 좌측 15도 각도에서 촬영한 얼굴영상을 이용할 경우, 어떤 알고리즘 기술이 표정과 각도 변화에서 강점을 가지는지 비교할 수 있다. 본 논문에서는 성능평가 결과에 영향을 줄 수 있는 항목들을 테스트항목이라고 말하고 성능평가 시스템 개발자는 평가 목적에 맞도록 테스트항목들을 설계하여야 한다. 생체인식 시스템에 있어서 성능에 영향을 끼치는 요인들에 대하여 JTC 1/Sc 37/WG 5 Part 1 표준안 [7]에서는 다음과 같이 분류하고 있다.

<표 1>에서 보이는 바와 같이 성능에 영향을 끼치는 요인들이 많으며, 이러한 요인들은 생체정보 인식시스템에 대한 성능평가를 수행할 때 프로브와 갤러리 이미지를 선정하는데 있어서 신중하게 고려해야 될 사항이다. 일반적으로 생체정보 데이터베이스를 구축할 때 위의 영향 요인들을 이용하여 생체정보 분류기준을 작성한다. 따라서 참조모델에서 선정되는 테스트항목은 생체정보 데

<표 1> 생체인식시스템의 성능에 영향을 끼치는 요인 분류

분류	세부요인
인구 통계적(Population demographics) 요인	나이, 민족 기원(ethnic origin), 성별, 직업 등
적용(Application) 요인	template ageing, 촬영시간, 사용자 친숙도, 사용자 동기 등
사용자 생리적(User physiology) 요인	수염, 대머리, 병과 질환, 속눈썹, 손톱길이, 키, 피부색 등
사용자 행동(User behavior) 요인	표정, 방언, 동작, 자세, 각도, 거리, 스트레스, 긴장도 등
환경적 영향(Environmental influences) 요인	배경(색, 그림자, 소리 등), 조명 강도 및 방향, 날씨(온도, 습도, 비, 눈 등)
사용자 외모(User appearance) 요인	복장(모자, 소매, 바지, 구두 등), 밴드(붕대)부착, 콘택트렌즈, 화장품, 안경부착, 가짜손톱, 헤어스타일, 반지, 문신 등
센서와 하드웨어(Sensor and hardware) 요인	카메라렌즈의 오염, 초점, 센서 품질, 센서 변경, 전송 채널 등
사용자 인터페이스(User Interface) 요인	피드백, 지시, 감시감독 여부

이터베이스의 분류기준 내로 한정된다.

3.1.2 실행모듈

실행모듈은 평가대상이 되는 생체정보 인식시스템을 동작시켜 성능평가를 실행한다. 성능평가 시스템과 생체정보 인식시스템과의 연동방식은 크게 두 가지가 있다. 첫 번째는 두 시스템을 독자적인 응용프로그램으로 개발하는 것이며, 두 번째는 생체정보 인식알고리즘을 컴포넌트나 라이브러리 형태로 작성한 후 이를 성능평가 시스템에 끼워 넣어 동작시키는 방식이다.

전자의 경우에는 성능평가에 이용될 입력 자료와 성능평가 수행결과 자료를 미리 정해진 양식-일반적으로 XML-의 파일 형태로 전달되기 때문에 입/출력 파일 양식에 대하여 두 시스템 간의 합의가 있어야 한다. FRVT 2006의 경우에는 성능평가 도구를 이용한 것은 아니지만, 업체들로부터 생체정보 인식시스템을 실행파일 형태로 넘겨 받아 평가에 이용되는 입력파일과 평가결과를 기록하는 출력파일 이름을 프로그램 실행 인자(argument)로 전달하는 방식으로 진행되었다.

컴포넌트(혹은 라이브러리) 방식의 경우에는 표준화된 인터페이스가 미리 합의되어야 한다. 합의된 인터페이스는 최대한 단순하여야 하며, 국제표준과 호환이 되는 것이 바람직하다. 관련 국제표준으로는 BioAPI 1.1과 BioAPI 2.0 등이 있다.

3.1.3 결과분석모듈

결과분석모듈은 실행모듈에서 수행한 결과값을 가지고 인식기술 알고리즘의 성능을 최종 분석한다. 특정 알고리즘 성능은 여러 가지 측정 항목 값으로 표현되며 성능평가 목적에 따라 측정 항목을 결정한다. 측정 항목은 크게 “에러비율(Error rates)” 과 “처리량비율(Throughput rates)”로 구분된다. 에러비율은 기본적으로 매칭에러(matching error)와 샘플획득에러(sample acquisition error)에 기인하며, 알고리즘이 제대로 정확히 동작하는가에 시험 초점을 맞추고 있다. 처리량비율은 얼

<표 2> JTC 1/Sc 37/WG 5 Part 1 표준안의 성능평가 메트릭

메트릭 분류	세부 메트릭
기본적 성능 메트릭	<ul style="list-style-type: none"> • 등록 실패율(Fail-to-enroll rate) • 획득 실패율(Fail-to-acquire rate) • 잘못된 비매칭률(False non-match rate) • 잘못된 매칭률(False match rate)
검증(Verification) 시스템 성능 메트릭	<ul style="list-style-type: none"> • 잘못된 거부율(False reject rate) • 잘못된 승인율(False accept rate) • 일반적 잘못된 거부율(General false reject rate) • 일반적 잘못된 승인율(General false accept rate)
식별(Identification) 시스템 성능 메트릭	<ul style="list-style-type: none"> • 식별률(Identification rate) • 잘못된 부정률(False-negative rate) • 잘못된 긍정률(False-positive rate)
닫힌집합 식별(Closed set identification)	<ul style="list-style-type: none"> • CMC 커브(curve)
DET/ ROC 커브	<ul style="list-style-type: none"> • ROC(Receiver Operating Characteristic) 커브 • DET(Detection Error Trade-off) 커브

굴인식 시스템이 단위시간 동안에 처리를 할 수 있는 사용자의 양을 보여준다. 이 처리량비율은 대용량 이미지 데이터베이스에서 검증(Verification) 작업을 수행할 때 특히 중요한 의미를 가진다.

다음 <표 2>는 JTC 1/Sc 37/WG 5 Part 1 표준안[7]에서 제안한 성능평가 측정 항목인 메트릭(Metric)들이다.

<표 2>에 나열된 측정 항목 외에도 처리량비율 관련해서 “평균 등록(매칭) 시간”, “평균(최대) 템플릿 크기”, “최대 등록(매칭) 사용 메모리” 등이 있다.

성능평가 도구는 선정된 측정 항목들의 값을 계산한 후 성능평가 결과보고서를 생성하여야 하며, 결과보고서에는 측정 항목 결과 값 외에도 성능평가를 수행한 환경과 조건에 관한 설명이 기술되어야 한다.

3.2 평가시스템 구축 프로세스

참조모델에 따라 성능평가 시스템을 구축하는

과정은 다음과 같다.

- (1) 성능평가 시스템의 구축 및 평가 목적을 기술한다.
- (2) 성능평가에 이용될 생체정보 데이터베이스를 구축하거나 선정한다.
- (3) 평가목적에 맞도록 테스트항목을 설계한다.
- (4) 성능평가 시스템과 생체정보 인식시스템과의 연동방식을 결정한다. 개별적인 프로그램으로 동작할 경우에는 입/출력 파일 양식을 결정하고, 컴포넌트 방식으로 연동할 경우에는 컴포넌트 인터페이스를 설계한다.
- (5) 평가목적에 맞도록 측정항목을 선정한다.
- (6) 생체정보 데이터베이스와 연동하여 테스트 항목에 따라 생체정보를 읽어오는 “자료준비모듈”을 구현한다.
- (7) 자료준비모듈에서 제공하는 갤러리/프로브 생체정보를 가지고 인식 알고리즘을 수행하는 “실행모듈”을 구현한다. 수행된 결과 값(유사도 값)은 성능평가결과 데이터베이스에 보관된다.
- (8) 성능평가결과 데이터베이스에 보관된 유사도 값을 분석하여 측정항목 값을 계산하고, 성능평가 결과보고서를 생성하는 “결과분석모듈”을 구현한다.

테스트항목과 측정 항목을 평가시스템 구축 시에 결정하지 않고, 실제 성능평가를 수행할 때 결정할 수도 있다. 이 경우에는 위 (3), (5) 단계에서 선택 가능한 테스트항목과 측정 항목들을 선정하고, 시험자가 성능평가 도구를 실행할 때 이 항목들 중에서 필요한 항목들을 선택할 수 있도록 구현되어야 한다.

4. 성능평가 도구 설계 및 구현

본 논문에서 제안된 참조모델을 이용하여 성능평가 도구를 설계하고 구현하였다. 평가시스템 구축 프로세스에 따른 단계별 내용과 그 결과물은

다음과 같다.

4.1 시스템 구축 및 평가 목적

성능평가 목적은 국내 시장에 나와 있는 얼굴인식시스템들의 기술수준을 파악하고, 객관적인 성능평가와 공신력 있는 인증을 통하여 국내 얼굴인식 기술의 발전을 도모하고 얼굴인식 제품들의 국제 경쟁력을 높이기 위한 것이다.

4.2 성능평가에 이용될 생체정보 데이터베이스

한국정보보호진흥원에서 2002년부터 2004년 말까지 구축한 연구용 얼굴데이터베이스를 성능평가에 이용한다. 이 데이터베이스는 얼굴이미지를 다음 <표 3>의 기준에 따라 분류하고 있다.

4.3 테스트항목

평가시스템을 구축 시에 테스트항목을 미리 결정하지 않고, 실제 성능평가를 수행할 때 시험자

<표 3> 연구용 얼굴데이터베이스 분류기준

항목	내용
성별	남자 혹은 여자
나이	나이
출생지	서울시, 경기도, ...
조명색	백색, 주황색
조명방향	정면, 우측/좌측 45도, 우측/좌측 90도, 우측/좌측 135도, 180도
안경착용방향	정면, 우측/좌측 45도, 우측/좌측 90도
얼굴표정	무표정, 웃는표정, 화난표정, 눈감은표정, 놀란표정
자세	정면, 우측/좌측 15도, 우측/좌측 30도, 우측/좌측 45도
헤어상태	자연스런 모발, 헤어밴드착용, 안경착용
촬영회차	1회, 2회, 3회
파일종류	원본영상, 압축영상, 요약영상 (Thumbnail)

가 선택할 수 있도록 테스트항목을 설계하였다. 기본적으로 테스트항목은 <표 3>의 분류기준을 모두 나열하여 시험자가 갤러리 영상 집합의 조건과 프로브 영상 집합의 조건으로 별도로 선택할 수 있도록 하였다. 이렇게 선택된 여러 항목의 갤러리 영상 집합과 여러 항목의 프로브 영상 집합을 테스트 집합 (Test-set)이라고 말한다. 하나의 성능평가 프로젝트는 테스트 집합을 여러 개 생성할 수 있으며, 각각의 테스트 집합마다 각기 다른 결과보고서를 생성할 수 있다.

예를 들어, 얼굴인식시스템에 등록할 갤러리 얼굴영상 사진의 집합이 백색-정면("Normal") 또는 주황색-정면("Illum, Yellow") 사진이고, 실제로 사용자들의 신원 인증을 위해 영상 획득 장비로 획득한 얼굴영상은 정상적이거나 눈을 감는다거나 왼쪽, 오른쪽으로 약간 틀어진 사진이 있을 수 있다. 이런 얼굴인식시스템을 가정 하여 테스트 집합을 구성하면 다음과 같다.

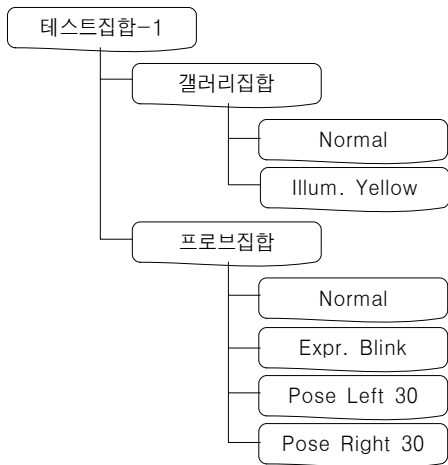
위의 예에서 만약 하나의 카테고리에 10명의 얼굴 영상이 존재한다고 가정하면, 갤러리 집합은 총 20개의 얼굴 영상이, 프로브 집합은 총 40개의 얼굴 영상이 존재하게 된다. 따라서 위 테스트집합에 대한 성능평가를 수행하면 갤러리 영상에 대해 20번의 템플릿 생성 함수가 호출되고, 프로브

영상에 대해 40번의 영상 처리가 이루어진 다음, 총 $20 \times 40 = 800$ 번의 비교연산이 수행되게 된다. 이 가운데 특정 인물에 대한 영상이 갤러리 집합에서 2번, 프로브 집합에서 4번 나오므로 동일 인물의 영상 비교는 한 사람 당 8번이 되고, 전체적으로는 총 10명이므로 80번이 된다. 따라서 서로 다른 인물간의 영상 비교는 720번이 된다. 이러한 방법으로 영상 비교 횟수를 미리 예상할 수 있으며, 비교 횟수를 미리 계산하여 보여줌으로써 시험자가 성능시험에 소요되는 시간을 미리 예상할 수 있도록 하였다.

4.4 인식시스템과의 연동방식

성능평가 도구에서는 얼굴인식을 위한 표준 인터페이스를 제공하고, 이 인터페이스를 만족하는 얼굴인식 모듈을 피시험자가 동적라이브리리 형태로 제공하도록 하였다. 성능평가 도구에서는 얼굴인식 모듈을 실행 도중(run-time)에 시험자의 선택에 따라 교체가 가능하도록 설계함으로써 성능평가 도구의 수정 없이 얼굴인식 모듈을 변경하여 성능평가를 수행할 수 있다.

국제표준과의 호환을 위하여 BioAPI 1.1 인터페이스를 준용하였으며, 피시험자의 개발 부담을 덜어주기 위하여 알고리즘 성능평가에 필요한 최



[그림 2] 얼굴영상 테스트집합 설정 예

<표 4> 성능평가 피시험 얼굴인식 모듈 인터페이스

함수명	의미
BioAPI_Init	초기화
BioAPI_Terminate	종료
BioAPI_FreeBIRHandle	BIR데이터 메모리 해제
BioAPI_GetBIRFromHandle	데이터 핸들로부터 BIR데이터 반환
BioAPI_CreateTemplate	갤러리로부터 데이터 템플릿 생성
BioAPI_Process	프로브 영상을 처리하여 BIR데이터 생성
BioAPI_VerifyMatch	처리된 두 영상을 비교하여 인증결과 반환

소한의 함수만 선정하였다. 연동을 위한 얼굴인식 모듈 인터페이스는 위의 <표 4>와 같다.

얼굴인식 모듈에서 제공하는 함수들은 성능평가 도구에 의하여 호출되는데, 그 주요한 호출흐름은 [그림 3]과 같다. 우선 BioAPI_Init 함수를 호출하여 얼굴인식 모듈을 초기화 한다. 그런 다음 성능 시험 과정에서 선택된 갤러리 이미지 각각에 대하여 BioAPI_CreateTemplate 함수를 호출하여 시스템에 등록하기 위한 템플릿 데이터를 생성한다. 그런 다음 선택된 프로브 이미지 각각에 대하여 BioAPI_Process를 호출하여 프로브 영상을 처리한 BIR 데이터를 생성한 다음 갤러리 영상으로부터 생성한 템플릿데이터와 함께 BioAPI_VerifyMatch를 호출하여 둘 간의 유사도를 계산한다. 유사도 계산은 프로브 이미지 데이터 각각에 대해 이미 생성되어 있는 모든 갤러리의 템플릿 데이터에 대해 반복한다. 이러한 과정을 모두 마치면 BioAPI_Terminate을 호출하여 종료한다.

4.5 측정 항목

FERET과 FRVT에서 사용된 성능평가 메트릭들과 JTC 1/Sc 37/WG 5 Part 1 표준안[7]에서 제안한 메트릭들을 분석하였다. 이들 중 얼굴인식 시스템에 대한 기술 평가(Technology evaluation)에 관련된 항목들을 선정하였으며, 선정된 메트릭은 다음과 같다.

<표 5> 선정된 성능평가 메트릭

<ul style="list-style-type: none"> • 등록 실패율(FTER) • 획득 실패율(FTAR) • 잘못된 거부율(FRR) = $F_{TA} + F_{NMR} * (1 - F_{TA})$ • 잘못된 승인율(FAR) = $F_{MR} * (1 - F_{TA})$ • 잘못된 비매칭률(FNMR) • 잘못된 매칭률(FMR) • CMC 커브 • ROC 커브 • 에러 교차율(Equal Error Rate) 	<ul style="list-style-type: none"> • 등록 처리량 (#건/sec.) • 등록 시간 (최소, 최대, 평균) • 추출 처리량 (#건/sec.) • 추출 시간 (최소, 최대, 평균) • 매칭 처리량 (#건/sec.) • 매칭 시간 (최소, 최대, 평균) • 트랜잭션(Transaction) 처리량 (#건/sec.) • 트랜잭션(Transaction) 시간 (최소, 최대, 평균) • 템플릿 크기(최소, 최대, 평균)
--	--

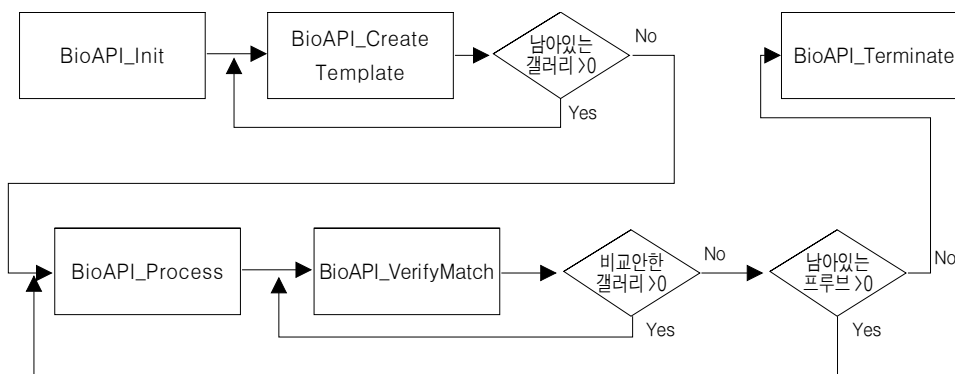
<표 5>에 선정된 메트릭에 의해 계산된 측정 항목 값은 평가 결과보고서에 출력된다.

4.6 자료준비모듈/실행모듈/결과분석 모듈 구현

성능평가 도구는 윈도우 운영체제의 응용프로그램으로 개발되었으며, 평가대상이 되는 얼굴인식 모듈은 동적라이브러리(Dynamic Link Library)로 구현된다.

성능평가에 이용될 생체정보 데이터베이스와의 연동기능과, 갤러리와 프로브 영상집합 설정기능을 제공하는 “자료준비모듈”을 구현한다.

업체에서 제공한 얼굴인식모듈이 <표 4>에서



[그림 3] 얼굴인식 모듈 인터페이스 호출 과정

제시한 함수들을 모두 제공하는지 여부를 검사하고, 선택된 얼굴인식모듈 기능을 이용하여 성능평가를 수행하는 “실행모듈”을 구현한다. 성능평가가 제대로 진행되고 있는지 여부를 시각적으로 보여주는 기능을 “실행모듈”에 포함시켰다.

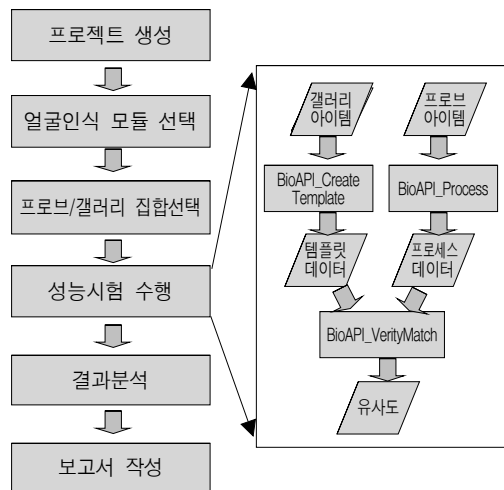
성능평가결과 데이터베이스에 보관된 유사도 값을 분석하여 측정항목 값을 계산하고, 성능평가결과보고서를 생성하는 “결과분석모듈”을 구현한다. 본 평가도구에서는 평가결과 보고서 뿐만 아니라 한 단계 더 나아가 평가결과에 따라 얼굴인식모듈에 대한 인증서도 발급하는 기능을 가지고 있다.

5. 성능평가 도구 실행 예

[그림 4]는 성능평가 도구에서 성능평가를 수행하는 절차를 보여준다.

5.1 프로젝트 생성 및 얼굴인식 모듈 선택

인증을 거친 시험자는 성능평가 수행을 위하여 새로운 프로젝트를 생성한다. 프로젝트가 생성되면 성능평가 도구는 시험자에게 성능평가하고자 하는 얼굴인식 모듈을 선택하도록 한다. 얼굴인식



[그림 4] 얼굴인식시스템 성능평가 절차

모듈이 선택되면 성능평가 도구는 선택된 얼굴인식 모듈에 대하여 동적 링크(dynamic link)로 연동된다. 연동된 얼굴인식 모듈은 성능평가 도구에서 요구하는 얼굴인식 모듈 인터페이스 명세를 만족하는지 여부를 검사하며, 해당 모듈이 명세를 만족할 경우에만 성능평가를 진행한다.

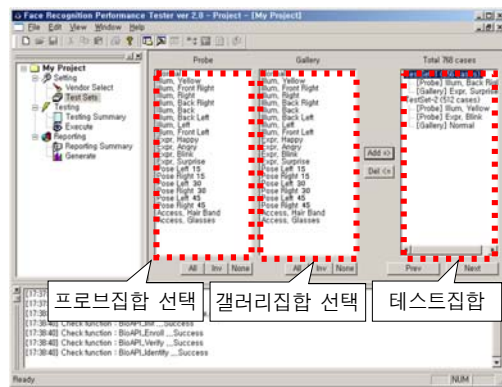
5.2 프로브/갤러리 집합 선택

그 다음으로 성능평가에 사용되는 얼굴영상 프로브 집합과 갤러리 집합을 선택한다. [그림 5]는 성능평가 도구에서 두 집합을 선택하는 모습을 보여준다.

성능평가 도구의 창에 보이는 얼굴데이터베이스의 분류기준에서, 평가자는 성능평가 목적에 맞도록 프로브 집합과 갤러리집합을 선택하여 테스트집합을 선정한다.

5.3 성능시험 수행

얼굴영상 프로브 집합과 갤러리 집합 선택이 완료되면 본격적으로 얼굴인식모듈의 성능을 시험하게 된다. 이 단계는 가장 많은 시간이 소요되는 작업으로 [그림 6]에서 보이는 바와 같이 템플릿 데이터 생성, 프로세스 데이터 생성, 인증 매칭의 세 가지 단계로 진행된다. 성능시험 단계에서 시



[그림 5] 얼굴영상 프로브 집합과 갤러리 집합 선택화면



[그림 6] 성능평가 시 진행 시각화 화면

험자는 성능평가 도구가 얼굴인식 모듈의 동작 과정을 직관적으로 보여주기 위하여 시각화 기능을 선택할 수 있다.

5.4 결과 분석

성능시험 단계에서 성능시험 도구는 하나의 프로브 이미지와 하나의 갤러리 이미지 간의 유사도 값과 인증 수행 시간 값의 두 가지 정보를 가지는 데이터 목록을 생성한다. 이러한 목록은 시험자가

선택한 테스트집합마다 생성되는데, 시험자는 특정 테스트집합을 선택하여 결과 보고서를 출력할 수 있다. 시험자가 보고서 출력을 원하는 테스트 집합을 선택하면, 성능 평가 도구는 선택된 테스트 집합에 대한 데이터 목록으로부터 보고서 출력을 위한 데이터를 생성한다. 생성되는 데이터에는 얼굴인식 모듈의 정확도 측면에서의 등록 실패율 (FTER), 획득 실패율(FTAR), FAR, FRR, 그리고 이 두 데이터 사이의 관계를 나타내는 ROC 그래프 및 ERR값, 그리고 CMC 그래프 등이 있다. 또한 얼굴인식 모듈의 속도 측면에서는 등록 영상 템플릿 추출 시간, 획득 영상 처리 시간, 인증(verification)을 위한 비교 시간, 각 트랜잭션에 걸리는 총 시간 등의 항목들이 있다. 이를 위해 각 데이터를 정렬한다거나 평균을 내는 등의 작업을 수행하여 결과를 성능평가 결과 데이터베이스에 저장한다.

5.5 평가 결과보고서 생성

이렇게 성능 평가를 수행하여 평가 결과보고서



(a) 보고서 첫 화면

시험결과

항목	측정값
FAR	0.0625 (15 fails/ 240 tries, when threshold = 90)
FRR	0.0625 (1 fails/ 16 tries, when threshold = 90)
EER	0.0625
Failure to Enroll Rate	0.0000 (0 fails/ 16 tries)
Failure to Acquire Rate	0.0000 (0 fails/ 16 tries)
Enrollment Throughput	5.46 (건/sec)
Enrollment Time	Min: 0.53/ Avg: 0.18/ Max: 0.11 (sec)
Extraction Throughput	5.46 (건/sec)
Extraction Time	Min: 0.53/ Avg: 0.18/ Max: 0.11 (sec)
Matching Throughput	7.49 (건/sec)
Matching Time	Min: 0.22/ Avg: 0.13/ Max: 0.09 (sec)
Transaction Throughput	3.15 (건/sec)
Transaction Time	Min: 0.75/ Avg: 0.32/ Max: 0.20 (sec)
Template Size	Min: 512/ Avg: 512/ Max: 512 (bytes)
Testing Start Date	2006-09-16 오전 5:29:28
Testing Finish Date	2006-09-16 오전 6:30:07

(b) 성능 측정 값(보고서에 포함)

[그림 7] 성능평가 결과보고서 예

를 생성할 수 있으며 생성된 보고서는 [그림 7]과 같은 화면으로 나타나며, 대략 5페이지 정도로 구성된다.

[그림 7]에 보이는 성능측정 값을 바탕으로 시험에 이용된 얼굴인식 모듈의 성능을 분석해보면, FAR과 FRR은 임계값 90에서 모두 0.0625가 측정되었다. EER은 공교롭게도 FAR과 FRR 두 값이 0.0625로서 같으므로 0.0625가 된다. 등록 및 획득 실패는 없으며 각 성능 평가 시간의 항목들도 잘 나와 있음을 알 수 있다.

시험자는 결과보고서에 나와 있는 시험 결과에 따라, 해당 얼굴인식 모듈에 대하여 얼굴인식 성능에 대한 인증서 발급여부를 결정할 수 있다. 만일 시험자가 성능평가 대상이 되는 얼굴인식 모듈을 인증한다면 인증서를 성능평가 도구 내에서 생성할 수 있다.

6. 성능평가 방법 비교

대표적인 얼굴인식 평가 사례인 FERET, FRVT

와 본 논문의 성능평가 도구를 이용한 평가방법을 비교하면 다음과 같다.

FERET과 FRVT와 같은 성능평가 프로그램에 비하여 본 논문에 제안된 성능평가 도구를 이용한 성능평가 방법은 다음과 같은 장점이 있다.

- 얼굴영상 데이터베이스의 유출을 근원적으로 차단한다.
- 국제표준에 부합하는 얼굴인식 모듈 개발을 독려한다.
- 성능평가 대상자와 성능평가 시험자를 분리한다.
- 별도의 성능결과 분석도구가 필요 없다.
- 평가비용을 대폭 줄일 수 있으며, 업체별로 개별 평가를 실행할 수 있다.

가장 최근에 실시된 FRVT 2006의 경우에는 평가자가 얼굴인식 모듈을 업체관계자로부터 넘겨받아 평가하여 얼굴 데이터베이스의 유출을 차단하며, 모든 업체관계자가 한 장소에 모일 필요 없이 시험이 진행되는 등, 기존 평가 방법의 문제

<표 7> FERET, FRVT와 제안된 성능평가 도구의 평가 방법 비교

	FERET	FRVT	제안된 평가도구
평가대상	• 얼굴인식기술 평가	• 얼굴인식기술 평가 • 시나리오 평가 • 운영 평가	• 얼굴인식기술 평가
얼굴DB 배포방법	• 현장에서 DB 배포(얼굴DB 유출가능)	• 현장에서 DB 배포(얼굴DB 유출가능) • FRVT 2006: 업체로부터 얼굴인식 모듈을 넘겨받음(얼굴DB 유출위험 없음)	• 업체로부터 얼굴인식 모듈을 넘겨받음 (얼굴DB 유출위험 없음)
성능평가 실행자	• 업체관계자가 실행	• 업체관계자가 실행 • FRVT 2006: 평가자가 실행	• 평가자가 실행
결과분석도구	• 성능결과 분석도구 필요	• 성능결과 분석도구 필요	• 별도의 분석도구 불 필요
결과보고서 생성	• 수작업	• 수작업	• 도구 내에서 자동 생성
개별 평가	• 불가능	• 불가능	• 업체별로 개별 평가 가능
평가 비용	• 모든 업체관계자가 한 장소에 모여야 됨(고비용)	• 모든 업체관계자가 한 장소에 모여야 됨(고비용) • FRVT 2006: 모일 필요 없음(저비용)	• 모일 필요 없음(저비용)
업체 부담	• 입/출력 파일 양식 맞춤	• 입/출력 파일 양식 맞춤	• 인식모듈 인터페이스 맞춤

점을 많이 해소하였다. 그러나, 별도의 성능결과 분석도구가 필요하며 결과보고서 생성은 수작업에 의존한다. 또한 FRVT 2006은 통합적인 성능평가 도구 없이, 평가환경 설정과 수행 그리고 결과분석을 일과성으로 진행되어 성능평가 프로그램의 재활용적인 측면에서 어려움이 있다.

7. 결 론

본 논문에서는 생체정보 인식시스템 성능평가를 위한 참조모델을 제안하였다. 제안된 참조모델은 관련 국제 표준안들과 호환이 되도록 설계되었으며, 이를 바탕으로 개발된 성능평가 도구의 일관성과 신뢰성 향상에 기여한다. 참조모델의 필요성은 다음과 같다.

- 성능 평가시스템 모델과 구축방법 제시
- 성능 평가시스템에 있어서 관련 국제표준을 준용케 함
- 성능 평가시스템의 일관성과 신뢰성을 향상 시킴
- 성능 평가시스템을 설계하고 구현하는데 가이드라인을 제시

참조모델을 바탕으로 얼굴인식시스템의 성능을 비교·평가할 수 있는 성능평가 도구를 설계 및 구현하였으며, 한국정보보호진흥원의 얼굴영상 데이터베이스와 국내 F사와 H사의 얼굴인식 제품을 대상으로 성능시험을 실제로 실시하였다. 하지만 본 논문에서 제안된 성능평가 도구가 제대로 활용되기 위해서는 여러 업체들의 적극적인 성능평가 참여가 필요하다. 이를 위하여 얼굴인식 시스템의 성능에 대한 인증서 발급 기준을 체계적으로 제시하고, 인증의 활성화를 이룰 수 있는 방안을 마련하는 것이 필요하다.

참 고 문 헌

[1] 김학일 외 10인, “ISO/IEC 국제표준기반의

BioAPI 표준적합성 시험도구 개선(v2.0)”, 한국정보보호진흥원, 기술표준연구 05-02 최종연구보고서, 2005.

[2] 문현준 외 3인, “얼굴인식시스템 성능시험 도구 개발”, 한국정보보호진흥원 기술표준연구 04-04 최종연구보고서, 2004.

[3] 변혜란 외 7인, “얼굴인식 시스템 성능평가 방법론 연구”, 한국정보보호진흥원 평가 1 연구 02-04 최종연구보고서, 2002.

[4] (주)위치비전, “연구용 얼굴 DB 확대 구축”, 한국정보보호진흥원 기술표준연구 04-01, 최종연구보고서 부록(A), 2004.

[5] Blackburn D. M., J. M. Bone, and P. J. Phillips, “FRVT 2000 Evaluation Report”, FRVT 2000 documents, February 2001.

[6] ISO/IEC JTC 1/SC 37, *Biometric Application Programming Interface-Part 1: BioAPI Specification*, ISO Standards, April 2006.

[7] ISO/IEC JTC 1/SC 37, *Biometric Performance Testing and Reporting-Part 1: Test Principles and Framework*, ISO Standards FCD, 2005.

[8] Grother P. J., R. J. Micheals, and P. J. Phillips, “Face Recognition Vendor Test 2002 Performance Metrics”, Proceedings 4th International Conference on Audio Visual Based Person Authentication, 2003.

[9] Phillips P. J., P. Rauss, and S. Der, “FERET (FacE REcognition Technology) Recognition Algorithm Development and Test Report”, ARL-TR-995, U. S. Army Research Laboratory, 1996.

[10] Phillips P. J., H. Wechsler, J. Huang, and P. Rauss, “The FERET Database and Evaluation Procedure for Face Recognition Algorithms”, *Image and Vision Computing Journal*, Vol.16, No.5(1998), pp.

- 295-306.
- [11] Phillips P. J., H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-recognition Algorithms", IEEE trans. Pattern Analysis and Machine Intelligence, Vol.22, No.10(2000).
- [12] Phillips P. J., A. Martin, C. L. Wilson, and M. Przybocki, "An Introduction to Evaluating Biometric Systems", IEEE Computer, February 2000, pp.56-63.
- [13] Phillips P. J., P. Grother, R. J. Micheals, D. M. Blackburn, E Tabassi, and J. M. Bone. "FRVT 2002: Overview and Summary", FRVT 2002 documents, March 2003.
- [14] Phillips P. J., P. Grother, R. J. Micheals, D. M. Blackburn, E Tabassi, and J. M. Bone. "FRVT 2002: Evaluation Report", FRVT 2002 documents, March 2003.
- [15] Phillips P. J., W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, M. Sharpe. "FRVT 2006 and ICE 2006 Large-Scale Results", NIST 7408, March 2007.
- [16] Rizvi S., P. J. Phillips, and H. Moon, "The FERET Verification Testing Protocol for Face Recognition Algorithms", Image and Vision Computing Journal, 1998.
- [17] The BioAPI Consortium, *BioAPI Specification Version 1.1*, March 2001.

◆ 저 자 소 개 ◆

**신 우 창 (wcshin@skuniv.ac.kr)**

서울대학교에서 전산학 학사·석사를 하고 동 대학에서 컴퓨터공학 박사
를 받았다. 2007년 현재 서경대학교 인터넷정보학과 교수로 재직 중이며
주요 연구 관심분야는 설계패턴, 컴포넌트기반 개발, 소프트웨어 재구조
화, 소프트웨어 아키텍처 등이다.

