

일반논문 (Regular Paper)

방송공학회논문지 제31권 제5호, 2026년 9월 (JBE Vol.31, No.5, September 2026)

<https://doi.org/10.5909/JBE.2026.31.5.1088>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

품질과 식별 유용성을 분리한 신뢰도 기반 얼굴·지문 인증 점수 융합

이 지 상^{a)}, 박 구 만^{b)*}

Reliability-Based Score Fusion for Face - Fingerprint Authentication by Separating Quality and Identification Utility

Jisang Lee^{a)} and Gooman Park^{b)*}

요 약

생체인증 시스템은 실제 사용 환경에서 조명 변화, 얼굴 가림, 움직임에 의한 흐림, 지문 접촉 불량, 건조 및 오염 등으로 입력 품질이 저하될 수 있다. 그러나 입력의 외형적 품질이 낮더라도 인증에 유용한 정보가 남아 있을 수 있으며, 반대로 외형적 품질이 양호한 입력도 실제 판정에는 충분히 기여하지 못할 수 있다. 본 연구는 얼굴·지문 복합생체인증에서 입력 품질점수(Q)와 식별 유용성점수(U)를 분리하여 산정하고, 두 점수를 결합한 신뢰도(R)를 이용해 얼굴과 지문 점수의 융합 가중치를 조정하는 방법을 제안한다. 실험에는 LFW와 CelebA 얼굴 데이터셋, FVC2004 DB1_B 지문 데이터셋을 사용하였다. 얼굴과 지문 데이터셋은 동일 사용자를 공유하지 않으므로, 실제 사용자 단위의 복합인증이 아니라 정답 유형을 맞춘 점수 수준 모의실험으로 평가하였다. 사후 분석용 식별 유용성 U_oracle은 정답 레이블을 이용하므로 실제 운용에는 사용할 수 없으며, 제안 구조의 분석적 상한을 확인하는 용도로만 사용하였다. 실제 운용 후보인 s_runtime은 정답 정보를 사용하지 않고 품질점수와 판정 기준값으로부터 계산한 판정 확신도를 이용하였다. 실험 결과 품질점수와 식별 유용성점수의 선형 상관관계는 대부분 절댓값 0.07 이하로 낮게 나타났다. CelebA+FVC에서 s_oracle은 AUC 0.585, EER 0.429, TAR@FAR 5% 0.161을 보였으나, LFW+FVC에서는 품질 기반 융합(s_q)의 AUC 0.587이 s_oracle의 0.575보다 높았다. 운영용 s_runtime은 AUC/EER에서 일관된 우위를 보이지 않았지만 두 조합 모두에서 가장 낮은 ECE를 나타냈다. 따라서 본 결과는 특정 융합 방식의 절대적 우수성보다는 품질과 판정 기여도를 분리해 해석하는 구조의 가능성과 한계를 보여주는 검증으로 해석한다.

Abstract

Biometric input quality can degrade in real operating environments due to illumination changes, facial occlusion, motion blur, poor fingerprint contact, dryness, and contamination. This study separately estimates input quality (Q) and identification utility (U), and uses a combined reliability score (R) to adjust face and fingerprint fusion weights. LFW and CelebA were used for face experiments and FVC2004 DB1_B for fingerprint experiments. Because the face and fingerprint datasets do not share the same subjects, the evaluation is a score-level simulation rather than a true subject-linked multimodal authentication experiment. The oracle utility U_oracle uses ground-truth labels and is therefore treated only as a post-hoc analytical upper bound. The operational method s_runtime replaces ground-truth utility with a decision confidence calculated from the distance to the validation threshold. The linear correlation between Q and U was generally weak, with absolute coefficients mostly below 0.07. In CelebA+FVC, s_oracle achieved an AUC of 0.585, EER of 0.429, and TAR@FAR 5% of 0.161, whereas in LFW+FVC the quality-based method s_q achieved a higher AUC (0.587) than s_oracle (0.575). The operational method did not consistently improve AUC or EER, but showed the lowest ECE in both fusion combinations. These results should be interpreted as an initial validation of the proposed reliability interpretation framework rather than evidence of universal performance superiority.

Keyword : Multimodal biometric authentication, Face - fingerprint score-level fusion, Sample quality assessment, Identification utility estimation, Reliability-based weight optimization

Copyright © 2026 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

1. 서론

1. 연구 배경

생체인증 기술은 개인 식별과 사용자 인증을 위한 핵심 수단으로 다양한 분야에서 활용되고 있다. 특히 얼굴인증과 지문인증은 센서 보급성이 높고 사용자가 익숙하게 사용할 수 있어 출입통제, 금융 인증, 모바일 단말 보안, 무인 서비스 등에 널리 적용된다. 그러나 실제 운용 환경에서는 입력 조건이 일정하지 않다. 얼굴영상은 조명 변화, 역광, 비정면 촬영, 움직임에 의한 흐림, 마스크나 안경에 의한 가림의 영향을 받으며, 지문영상은 손가락의 건조 또는 과습, 부분 접촉, 압력 차이, 오염, 위치 치우침의 영향을 받는다. 이러한 입력 품질 저하는 특징 추출을 불안정하게 만들고 인증 오류를 증가시킨다.

이 문제를 줄이기 위해 두 가지 이상의 생체정보를 결합하는 복합생체인증이 활용된다. 얼굴과 지문은 획득 방식과 특징 구조가 다르므로 한쪽 입력의 상태가 나빠더라도 다른 입력이 전체 판정을 보완할 수 있다. 하지만 모든 입력을 같은 비중으로 결합하면 품질이 낮거나 실제 식별에 도움이 적은 입력이 최종 판정에 부정적 영향을 줄 수 있다. 따라서 각 입력을 얼마나 믿을 수 있는지 해석하고 그 신뢰 수준을 융합 가중치에 반영하는 구조가 필요하다.

2. 품질점수만 사용하는 방식의 한계

기존 생체인증 연구에서는 입력 영상의 상태를 하나의

품질점수로 요약하여 샘플 수용 여부나 융합 가중치에 활용하는 경우가 많다. 그러나 입력의 외형적 품질이 낮다는 사실이 곧바로 식별에 도움이 되지 않는다는 뜻은 아니다. 얼굴의 일부가 가려져 있어도 핵심 식별 영역이 유지될 수 있고, 지문 일부가 손상되어도 유효 특징점이 충분히 남을 수 있다. 반대로 품질이 좋아 보이는 입력도 동일 대상과 다른 대상을 충분히 구분하지 못할 수 있다. 따라서 품질점수만으로 신뢰도를 결정하면 저품질-고유용성 입력을 불필요하게 낮게 평가하거나 고품질-저유용성 입력에 과도한 가중치를 부여할 수 있다.

3. 연구 목적 및 기여

본 연구의 목적은 얼굴 및 지문 입력에 대해 품질점수(Q)와 식별 유용성점수(U)를 별도로 산정하고, 두 정보를 결합한 신뢰도(R)를 이용해 복합생체인증의 점수 융합을 해석 가능한 방식으로 조정하는 데 있다. 본 논문은 Q와 U의 도출식, 사후 분석용 *oracle*과 실제 운용 가능한 *runtime*의 계산 절차, 그리고 최종 가중치 계산식을 명시하여 재현성을 높인다.

- 얼굴과 지문 입력의 외형적 품질 Q와 판정 기여도 U를 구분하는 신뢰도 구조를 제안한다.
- 정답 레이블을 사용하는 *U_oracle*과 정답 없이 계산 가능한 운영용 판정 확신도 *C_runtime*을 명확히 구분한다.
- Q, U 또는 C를 결합한 신뢰도 R을 모달리티별 융합 가중치로 정규화하는 계산 절차를 제시한다.
- 낮은 AUC와 작은 지문 표본, 단순 매치, 동일 사용자 미연결이라는 한계를 결과 해석에 직접 반영한다.

다만 본 연구에서 사용한 얼굴·지문 데이터셋은 동일 사용자를 공유하지 않으므로, 본 실험은 두 모달리티의 비교 쌍 점수를 정답 유형에 맞추어 결합한 점수 기반(*chimeric*) 모의실험이다. 따라서 실제 사용자 단위 복합인증 성능을 직접 입증하는 실험이 아니라 제안 절차의 타당성을 확인하기 위한 1차 검증으로 해석한다.

a) 서울과학기술대학교 정보통신미디어공학과(Seoul National University of Science and Technology)

b) 서울과학기술대학교 스마트ICT융합공학과(Seoul National University of Science and Technology)

✉ Corresponding Author : 박구만(Gooman Park)

E-mail: gmpark@seoultech.ac.kr

Tel: +82-2-970-6430

ORCID: <https://orcid.org/0000-0002-7055-5568>

· Manuscript received August 12, 2026; Revised August 24, 2026;

Accepted August 24, 2026.

4. 주요 용어 정리

표 1. 본 논문에서 사용하는 주요 용어의 쉬운 설명

Table 1. Descriptive explanations of the key terms used in this paper

Conventional Term	Descriptive Expression in This Paper	Meaning
genuine pair	same-subject comparison pair	A pair formed by comparing two inputs obtained from the same person (or the same finger)
impostor pair	different-subject comparison pair	A pair formed by comparing two inputs obtained from different people (or different fingers)
score_s	matching similarity score	A score from 0 to 1 indicating how similarly the two inputs resemble the same subject
Q	quality score	A score indicating how good the input is in terms of brightness, sharpness, contact condition, etc.
U	identification utility score	A score indicating how much the input helps distinguish the same subject from a different subject
R	reliability (confidence)	The final confidence level obtained by combining the quality score with the utility score or decision confidence
oracle	post-hoc (oracle) analysis	Analysis conducted with knowledge of the ground truth, used to observe the upper-bound effect of the proposed method
runtime	operational (runtime)	A method computable without knowledge of the ground truth, as in an actual authentication process
FAR/FRR/TAR/EER	False Acceptance Rate / False Rejection Rate / True Acceptance Rate / Equal Error Rate	Key metrics used to evaluate biometric authentication performance
ECE	Expected Calibration Error	A metric measuring how well the confidence reported by the system matches the actual correctness rate

II. 관련 연구

1. 얼굴·지문 품질 평가 연구

지문인증 분야에서 품질 평가는 능선과 골의 선명도, 명암 대비, 능선의 연속성, 지문 중심부 포함 여부, 오염 및 부분접촉 여부 등을 이용하여 입력이 매칭에 적합한 상태인지 판단하는 방식으로 발전해 왔다. 얼굴인증 분야에서도 조명, 흐림, 포즈, 가림, 해상도, 얼굴 기준점 검출 안정성 등이 품질 평가에 활용된다. 이러한 품질 평가는 저품질 입력으로 인한 인증 실패를 줄이는 데 기여하지만, 입력이 실제 동일 대상과 다른 대상을 구분하는 데 얼마나 기여하는지는 별도의 문제이다.

2. 복합생체인증과 신뢰도 기반 융합

복합생체인증은 여러 생체정보를 함께 사용하여 단일 생체정보의 한계를 보완하는 방법이다. 특징 수준, 점수 수준, 결정 수준 융합 가운데 본 연구는 각 모달리티의 매칭 점수를 결합하는 점수 수준 융합을 대상으로 한다. 점수 수준 융합은 구현이 비교적 단순하지만 입력 상태에 따라 적절한 가중치를 부여하지 않으면 신뢰도가 낮은 입력의 영향을 충분히 줄이기 어렵다.

본 연구는 품질점수뿐 아니라 판정 기준값과의 관계에서 얻는 식별 유용성 또는 판정 확신도를 별도 정보로 취급한다. 특히 정답 정보를 이용하는 `s_oracle`은 실제 운용 방식이 아니라 분석용 상한으로만 사용하고, 실제 운용 가능성은 정답 정보를 사용하지 않는 `s_runtime`으로 별도 평가한다. 이 구분은 사후 분석 성능과 실제 운용 가능성을 혼동하지 않기 위한 것이다.

III. 제안 방법

1. 전체 절차

제안 방법은 입력 획득, 전처리 및 매칭, 품질점수 계산, 식별 유용성 또는 판정 확신도 계산, 신뢰도 계산, 최종 점수 융합의 순서로 구성된다. 얼굴 입력은 정렬 및 크기 조정 후 매칭 점수를 계산하고, 지문 입력은 지문 영역 보정과 선명도 향상 후 특징점 기반 점수를 계산한다. 이후 각 모달

리티의 신뢰도를 산정하고 신뢰도가 높은 모달리티에 더 큰 융합 가중치를 부여한다. 전체 흐름은 그림 1과 같다.

2. 품질점수 Q 산정

품질점수 Q는 입력이 물리적·영상적으로 얼마나 양호한지를 나타내는 0~1 범위의 값이다. 얼굴 입력에서는 밝기 적정성, 명암 대비, 흐림(초점) 정도를 사용하고, 지문 입력에서는 명암 대비, 선명도, 유효 지문 영역 비율을 사용한

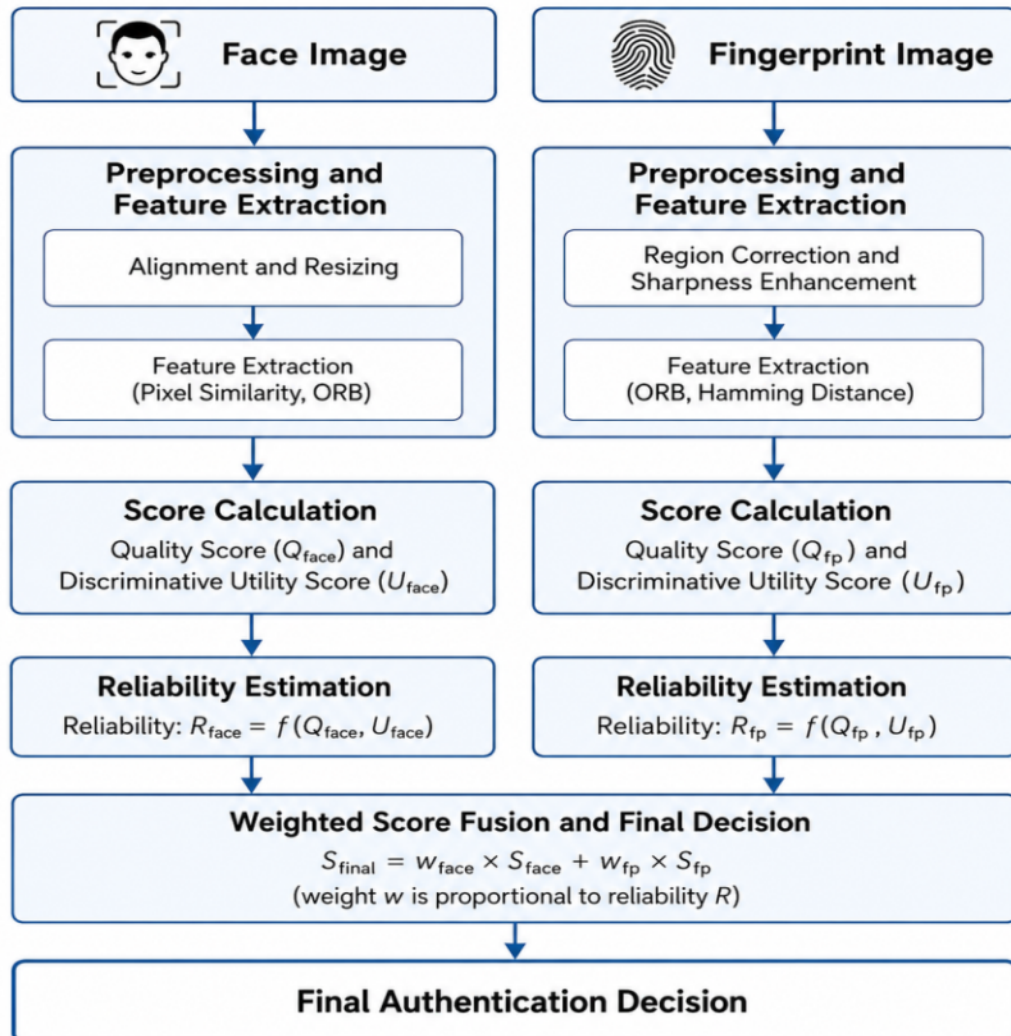


그림 1. 전체 절차 도식도

Fig. 1. Overall block diagram of the proposed method

다. 각 구성요소는 0~1로 정규화한 뒤 동일 가중 평균으로 단일 입력 품질을 계산하였다.

$$q_face(x) = [q_brightness(x) + q_contrast(x) + q_focus(x)] / 3$$

$$q_finger(x) = [q_contrast(x) + q_sharpness(x) + q_coverage(x)] / 3$$

비교쌍 i 가 두 입력 $x_{i,1}$ 과 $x_{i,2}$ 로 구성될 때 모달리티 m 의 비교쌍 품질점수는 두 단일 입력 품질의 평균으로 정의한다.

$$Q_m(i) = \{ q_m(x_{i,1}) + q_m(x_{i,2}) \} / 2, \quad m \in \{face, finger\}$$

따라서 Q 는 정답 레이블과 무관하게 영상 자체의 외형적 상태만을 표현하며, 1에 가까울수록 품질이 양호한 입력을 의미한다.

3. 식별 유용성점수 U 산정

식별 유용성점수 U 는 해당 비교쌍의 매칭 점수가 정답 방향으로 판정 기준값에서 얼마나 충분히 떨어져 있는지를 나타낸다. 여기서 $S_m(i)$ 는 모달리티 m 의 매칭 유사도 점수, T_m 은 검증용 데이터에서 산출한 판정 기준값, y_i 는 동일 대상(genuine)이면 1, 다른 대상(impostor)이면 0인 정답 레이블이다. 사후 분석용 식별 유용성은 다음과 같이 정의한다.

$$U_oracle, m(i) = \sigma \{ \alpha \cdot (2y_i - 1) \cdot [S_m(i) - T_m] \}$$

$\sigma(\cdot)$ 는 값을 0~1 범위로 변환하는 시그모이드 함수이고 $\alpha > 0$ 은 판정 여유의 스케일을 조정하는 계수이다. α 와 T_m 은 검증용 데이터에서 고정하고 시험 데이터의 정답을 이용해 조정하지 않는다. genuine의 경우 $S_m > T_m$ 일수록 U_oracle 이 커지고, impostor의 경우 $S_m < T_m$ 일수록 U_oracle 이 커진다.

중요하게도 U_oracle 은 y_i 를 직접 사용하므로 판정 성공 여부와 높은 관련성을 갖는 것은 정의상 어느 정도 예상된다. 따라서 U_oracle 과 정확도 간 관련성 자체를 실제 운용 성능의 증거로 해석하지 않고, 제안 구조가 이상적으로 작동할

때의 분석용 상한을 확인하는 진단 변수로만 사용한다.

실제 인증 단계에서는 정답을 알 수 없으므로 U_oracle 을 사용할 수 없다. 운영용 방식에서는 점수가 기준값에서 얼마나 떨어져 있는지만으로 판정 확신도 $C_runtime$ 을 계산한다.

$$C_runtime, m(i) = \sigma \{ \alpha \cdot |S_m(i) - T_m| \}$$

$C_runtime$ 은 판정 방향의 정답 여부는 알지 못하지만 기준값에서 멀리 떨어진 점수를 더 확신도 높은 판정으로 취급한다. 이 때문에 $s_runtime$ 은 실제 계산 가능한 방식이지만 s_oracle 과 동일한 성능을 기대할 수는 없다.

4. 신뢰도 R 과 점수 융합

본 1차 실험에서는 품질과 유용성 정보를 동일 비중으로 결합하기 위해 $\beta=0.5$ 를 사용하였다. 사후 분석용 신뢰도와 운영용 신뢰도는 각각 다음과 같이 계산한다.

$$R_oracle, m = \beta \cdot Q_m + (1 - \beta) \cdot U_oracle, m, \quad \beta = 0.5$$

$$R_runtime, m = \beta \cdot Q_m + (1 - \beta) \cdot C_runtime, m, \quad \beta = 0.5$$

얼굴과 지문의 최종 융합 가중치는 두 신뢰도의 상대적 비율로 정규화한다. $\varepsilon=10^{-8}$ 은 분모가 0이 되는 경우를 방지하기 위한 작은 상수이다.

$$w_face = R_face / (R_face + R_finger + \varepsilon)$$

$$w_finger = R_finger / (R_face + R_finger + \varepsilon)$$

$$S_final = w_face \cdot S_face + w_finger \cdot S_finger$$

품질 기반 방식 s_q 는 위 식에서 R 대신 Q 를 사용하여 가중치를 계산하고, 단순 평균 s_avg 는 $w_face=w_finger=0.5$ 로 고정한다.

사후 분석용 oracle의 적용 절차는 다음과 같다.

- 1단계: 검증용 데이터에서 모달리티별 판정 기준값 T_face 와 T_finger 를 결정한다.
- 2단계: 시험 비교쌍에 대해 매칭 점수 S 와 품질점수

표 2. 융합 방식의 설명형 정의

Table 2. Descriptive definitions of the fusion methods

Method	Plain-Language Meaning	Feasibility in Real Operation
s_avg	A baseline method that averages the face score and the fingerprint score with equal weights	Feasible
s_q	A method that adjusts the weights according to the quality scores of the face and the fingerprint	Feasible
s_oracle	A post-hoc (oracle) method that additionally reflects the identification utility computed using ground-truth information	Not feasible
s_runtime	An operational (runtime) method that uses the quality score and the decision confidence without any ground-truth information	Feasible

Q를 계산한다.

- 3단계: 오프라인 분석에서만 정답 y 를 사용하여 U_{oracle} 을 계산한다.
- 4단계: R_{oracle} 과 모달리티별 가중치 w 를 계산하고 최종 융합 점수 S_{final} 을 산출한다.
- 5단계: oracle 결과는 분석용 상한으로만 보고하며 실제 시스템 배치에는 사용하지 않는다.

IV. 데이터셋 및 실험 환경

본 연구는 얼굴 데이터셋으로 LFW와 CelebA를 사용하고, 지문 데이터셋으로 FVC2004 DB1_B를 사용하였다. VGGFace2, CASIA-WebFace, NIST SD, PolyU HRF, SOCOFing 등은 후속 확장 실험 후보이며 본 실험의 성능 수치에는 포함하지 않았다. 얼굴과 지문 데이터셋은 동일 사용자를 공유하지 않으므로 실제 사용자 단위의 복합인증이 아니라 점수 수준 모의실험으로 수행하였다. 데이터셋 구성은 표 3과 같다.

표 3의 실제 분석 대상은 LFW, CelebA, FVC2004 DB1_B이며, 그림 2~4는 각 데이터셋의 입력 예를 보여준다.

표 3. 데이터셋 구성 및 용도

Table 3. Composition and purpose of the datasets

Modality	Dataset	Main Purpose	Remarks
Face	LFW	Face verification experiments	Construction of same-subject and different-subject comparison pairs
Face	CelebA	Face quality and attribute-based analysis	Use of aligned face images and attribute annotations
Fingerprint	FVC2004 DB1_B	Fingerprint matching and quality analysis	Repeated captures of the same finger
Fingerprint	SOCOFing	Experiments under subsequent damage and alteration conditions	To be used in future dataset expansion

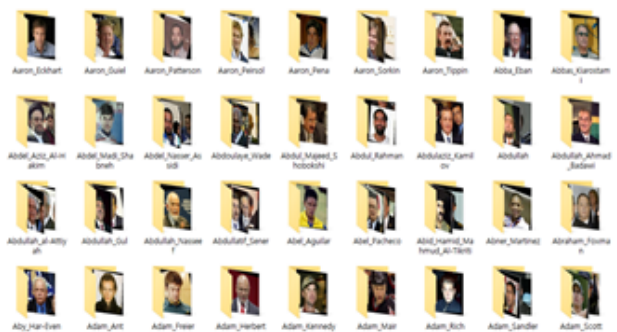


그림 2. LFW 데이터셋 예시

Fig. 2. Examples from the LFW dataset



그림 3. CelebA 데이터셋 예시

Fig. 3. Examples from the CelebA dataset



그림 4. FVC2004 DB1_B 지문 데이터셋 예시
Fig. 4. Examples from the FVC2004 DB1_B fingerprint dataset

1. 비교쌍 생성 및 데이터 분리

얼굴 데이터셋에서는 같은 인물의 서로 다른 이미지쌍을 동일 대상 비교쌍으로 구성하고, 서로 다른 인물의 이미지쌍을 다른 대상 비교쌍으로 구성하였다. 지문 데이터셋에서는 같은 손가락을 여러 번 촬영한 이미지쌍을 동일 대상 비교쌍으로 구성하고, 서로 다른 손가락 이미지쌍을 다른 대상 비교쌍으로 구성하였다. 데이터 중복에 의한 성능 과대평가를 줄이기 위해 먼저 인물 ID 또는 손가락 ID 기준으로 학습용, 검증용, 시험용 데이터를 분리한 뒤 각 데이터 안에서만 비교쌍을 생성하였다. 생성 결과는 표 4와 같다.

표 4. 비교쌍 생성 결과
Table 4. Results of comparison pair generation

Dataset	Training Pairs	Validation Pairs	Test Pairs	Remarks
LFW	174,478	289,000	21,036	Subject IDs derived from folder names
CelebA	259,838	56,216	55,946	Based on identity_CelebA.txt
FVC2004 DB1_B	336	112	112	Based on finger IDs

표 5. 평가 지표의 쉬운 설명
Table 5. Plain-language descriptions of the evaluation metrics

Metric	Description
Accuracy	The proportion of correctly classified comparison pairs among all pairs
False Acceptance Rate (FAR)	The proportion of different-subject pairs incorrectly accepted as same-subject pairs
False Rejection Rate (FRR)	The proportion of same-subject pairs incorrectly rejected as different-subject pairs
True Acceptance Rate (TAR)	The proportion of same-subject pairs correctly accepted
Equal Error Rate (EER)	The error rate at the operating point where the false acceptance rate and the false rejection rate become equal
AUC	A metric indicating how well the scores separate same-subject pairs from different-subject pairs
Expected Calibration Error (ECE)	The discrepancy between the confidence reported by the system and the actual correctness rate

표 4에서 얼굴 데이터는 비교적 큰 규모이지만 FVC2004 DB1_B는 시험용 112쌍으로 제한된다. 이 제한은 이후 얼굴·지문 융합 시험의 유효 표본 수와 낮은 FAR 영역 평가에 직접 영향을 준다.

2. 실험 설정 및 평가 지표

매칭 유사도 점수는 두 입력이 얼마나 같은 대상처럼 보이는지를 나타내는 0~1 범위의 값이다. 본 연구에서는 전체 Q-U-R 절차의 동작 가능성을 우선 확인하기 위해 얼굴에는 64×64 흑백 이미지의 단순 픽셀 유사도를, 지문에는 ORB 특징점과 해밍 거리 기반 유사도를 적용하였다. 이는 최신 생체인식 매처보다 판별력이 낮을 수 있으므로 본 연구의 AUC 0.5~0.6 수준을 상용 시스템의 성능으로 일반화해서는 안 된다. 특히 표 6에서 FVC의 동일/다른 대상 점수 분리도가 0.0069에 불과해 지문 매치의 제한이 융합 성능의 상한을 크게 낮추는 요인으로 작용한다.

낮은 AUC를 보완하여 해석하기 위해 기존 시험 결과에서 AUC와 EER뿐 아니라 TAR@FAR 5%, 1%, 0.1%, ECE, Brier score, AUC 95% 신뢰구간을 함께 제시한다. 다만 시험 impostor가 56쌍뿐이므로 FAR 1%와 0.1%는 안정적인

운영 성능으로 해석하지 않고 참고값으로만 사용한다.

V. 실험 결과

1. 품질점수와 매칭 유사도 점수 결과

품질점수는 데이터셋 특성과 대체로 일치하는 경향을 보였다. 정렬된 얼굴 이미지가 많은 CelebA는 평균 품질점수가 약 0.80으로 가장 높았고, LFW는 다양한 촬영 조건으로 인해 약 0.74~0.75 수준을 보였다. FVC2004 DB1_B는 약 0.71~0.76 수준이었다. 동일 대상 비교쌍과 다른 대상 비교쌍 간 품질 차이는 크지 않아 정답 유형이 Q 자체에 큰 편

향을 만들지 않았음을 확인하였다.

매칭 유사도는 모든 데이터셋에서 동일 대상 평균이 다른 대상 평균보다 높아 점수 방향은 정상적이었다. 그러나 표 6과 같이 차이는 매우 작았고, 특히 FVC2004 DB1_B의 분리도는 0.0069였다. 이는 이후 융합 AUC가 낮게 나타나는 핵심 원인 중 하나로 해석한다.

2. 품질점수와 식별 유용성점수의 관계

품질점수와 사후 분석용 식별 유용성점수의 상관분석 결과, 대부분의 데이터 구간에서 상관계수의 절댓값은 0.07 이하로 나타났다. 이는 두 점수가 통계적으로 독립임을 의미하는 것은 아니며, 본 실험 범위에서 강한 선형 관계가

표 6. 시험용 데이터 기준 매칭 유사도 점수의 분리 경향

Table 6. Separation trends of matching similarity scores on the test data

Dataset	Same-Subject Mean	Different-Subject Mean	Separation	Interpretation
LFW	0.5994	0.5648	0.0347	Same-subject scores are higher, but the separability is limited
CelebA	0.5921	0.5487	0.0434	Slightly stronger separation tendency than LFW
FVC2004 DB1_B	0.7096	0.7027	0.0069	Strongly limited by the fingerprint matcher and the small sample size

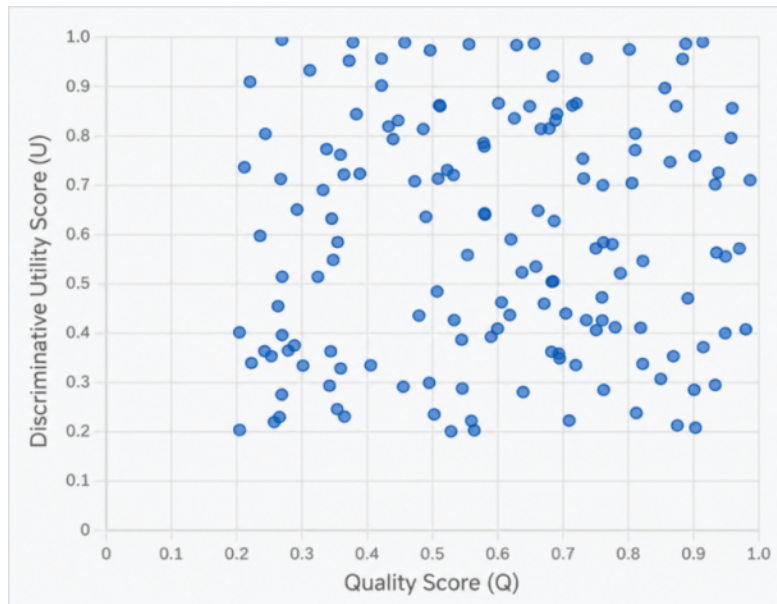


그림 5. 품질점수(Q)와 식별 유용성점수(U)의 산점도

Fig. 5. Scatter plot of the quality score (Q) versus the identification utility score (U)

표 7. 품질점수와 식별 유용성점수 관계에 대한 주요 관찰

Table 7. Key observations on the relationship between the quality score and the identification utility score

Analysis Item	Result	Implication
Relationship between the two scores	Correlation coefficients mostly below 0.07	Quality and identification utility represent different types of information
Low-quality, high-utility cases	Frequently observed in CelebA and LFW	CelebA and LFW Even low-quality inputs can contribute to authentication
High-quality, low-utility cases	Observed across all datasets	Even high-quality inputs can be disadvantageous for the decision
Utility score and decision success	A strong association was confirmed	The utility score is directly related to the actual decision outcome

관찰되지 않았다는 뜻이다. 실제로 저품질-고유용성 및 고 품질-저유용성 사례가 함께 확인되었다. 그림 5는 Q와 U의 산점도를, 표 7은 주요 관찰을 요약한다.

특히 표 7의 “유용성점수와 판정 성공 여부” 관계는 U_oracle이 정답 방향을 식에 포함하기 때문에 어느 정도 정의상 발생한다. 따라서 이 결과를 제안 방식의 실용적 성능 우수성으로 해석하지 않고, Q와 정답 방향 판정 여유가 서로 다른 정보를 표현하는지 확인하는 사후 진단 결과로 한정한다.

3. 신뢰도 및 얼굴·지문 융합 결과

얼굴과 지문 데이터셋이 동일 사용자를 공유하지 않으므

로, 본 연구는 얼굴 비교쌍 점수와 지문 비교쌍 점수를 정답 유형에 맞추어 결합하는 점수 기반 모의실험을 수행하였다. 표 8에서 모든 방식은 동일 대상 평균 점수가 다른 대상 평균보다 높았고, s_oracle은 두 조합 모두에서 단순 평균보다 큰 “평균 점수 분리도”를 보였다. 그러나 이 결과는 정답 정보를 사용하는 분석용 상한이며 실제 운영 성능으로 해석할 수 없다.

4. 융합 방식별 인증 성능

융합 성능은 검증용 데이터에서 산출한 동일오류율 기준 판정값을 시험용 데이터에 적용하여 평가하였다. 시험용 데이터는 각 융합 조합별 112개 비교쌍으로 제한되므로 절

표 8. 시험용 데이터 기준 융합 점수 분리도

Table 8. Separation of fusion scores on the test data

Fusion	s_avg	s_q	s_oracle	s_runtime	Main Interpretation
LFW+FVC	0.0197	0.0240	0.0393	0.0219	The oracle-reliability fusion yields the largest separation
CelebA+FVC	0.0214	0.0245	0.0514	0.0254	The oracle-reliability fusion shows the greatest improvement

표 9. 시험용 데이터 기준 융합 방식별 인증 성능

Table 9. Authentication performance of each fusion method on the test data

Fusion	Method	Acc	FAR	FRR	TAR	AUC	EER
CelebA+FVC	s_avg	0.571	0.250	0.607	0.393	0.551	0.446
CelebA+FVC	s_q	0.571	0.250	0.607	0.393	0.565	0.446
CelebA+FVC	s_oracle	0.562	0.250	0.625	0.375	0.585	0.429
CelebA+FVC	s_runtime	0.562	0.250	0.625	0.375	0.565	0.429
LFW+FVC	s_avg	0.527	0.286	0.661	0.339	0.569	0.420
LFW+FVC	s_q	0.545	0.268	0.643	0.357	0.587	0.411
LFW+FVC	s_oracle	0.536	0.304	0.625	0.375	0.575	0.446
LFW+FVC	s_runtime	0.527	0.268	0.679	0.321	0.574	0.420

대 성능보다 방식 간 상대적 경향을 중심으로 해석한다. 구체 결과는 표 9와 같다.

CelebA+FVC에서는 s_oracle 이 AUC 0.585, EER 0.429로 해당 조합 내 가장 높은 분석용 AUC를 보였다. 반면 LFW+FVC에서는 s_q 가 AUC 0.587, EER 0.411로 가장 양호했고 s_oracle 의 AUC는 0.575였다. 따라서 “oracle이 항상 단순/품질 기반 융합보다 우수하다”는 결론은 성립하지 않는다.

LFW+FVC에서 oracle이 s_q 보다 낮은 이유는 U_oracle 이 각 비교쌍의 정답 방향 판정 여유를 반영할 뿐, 전체 시험 점수의 순위를 직접 최적화하여 AUC를 최대화하는 목적함수가 아니기 때문이다. AUC는 전체 genuine/impostor 점수의 순위 관계에 의해 결정되므로, 분리력이 약한 FVC 점수와 112쌍의 작은 표본에서 비교쌍별 가중치가 변경되

면 일부 점수의 전역 순위가 오히려 불리하게 바뀔 수 있다. 즉 oracle은 “정답을 쓰면 반드시 AUC가 최댓값이 되는 최적화 알고리즘”이 아니라, 정답 기반 유용성을 가중치에 반영했을 때의 사후 분석 조건이다.

운영용 $s_runtime$ 은 CelebA+FVC에서 AUC 0.565로 s_q 와 동일했고, LFW+FVC에서 AUC 0.574로 s_q (0.587)보다 낮았다. 따라서 운영용 방식의 실질적 AUC 우위는 현재 실험에서 입증되지 않았다.

5. 정밀 평가 결과

낮은 AUC를 보완하여 결과를 다각도로 평가하기 위해 기존 시험 점수에서 TAR@FAR, ECE, Brier score와 AUC 95% 신뢰구간을 함께 확인하였다. 표 10에서 CelebA

표 10. 시험용 데이터 기준 TAR@FAR 및 신뢰도 보정 결과
Table 10. TAR@FAR and calibration results on the test data

Fusion	Method	TAR@FAR 5%	TAR@FAR 1%	TAR@FAR 0.1%	ECE	Brier
CelebA+FVC	s_avg	0.125	0.054	0.054	0.168	0.272
CelebA+FVC	s_q	0.089	0.071	0.071	0.159	0.269
CelebA+FVC	s_oracle	0.161	0.054	0.054	0.202	0.264
CelebA+FVC	$s_runtime$	0.107	0.054	0.054	0.146	0.269
LFW+FVC	s_avg	0.054	0.018	0.018	0.163	0.273
LFW+FVC	s_q	0.054	0.036	0.036	0.163	0.270
LFW+FVC	s_oracle	0.071	0.054	0.054	0.170	0.271
LFW+FVC	$s_runtime$	0.071	0.018	0.018	0.161	0.271

표 11. AUC 95% 신뢰구간 및 동일오류율
Table 11. AUC with 95% confidence intervals and equal error rates

Fusion	Method	AUC	95% CI	EER
CelebA+FVC	s_avg	0.551	[0.444, 0.658]	0.446
CelebA+FVC	s_q	0.565	[0.467, 0.673]	0.446
CelebA+FVC	s_oracle	0.585	[0.483, 0.691]	0.429
CelebA+FVC	$s_runtime$	0.565	[0.455, 0.668]	0.429
LFW+FVC	s_avg	0.569	[0.464, 0.678]	0.420
LFW+FVC	s_q	0.587	[0.479, 0.701]	0.411
LFW+FVC	s_oracle	0.575	[0.469, 0.685]	0.446
LFW+FVC	$s_runtime$	0.574	[0.470, 0.680]	0.420

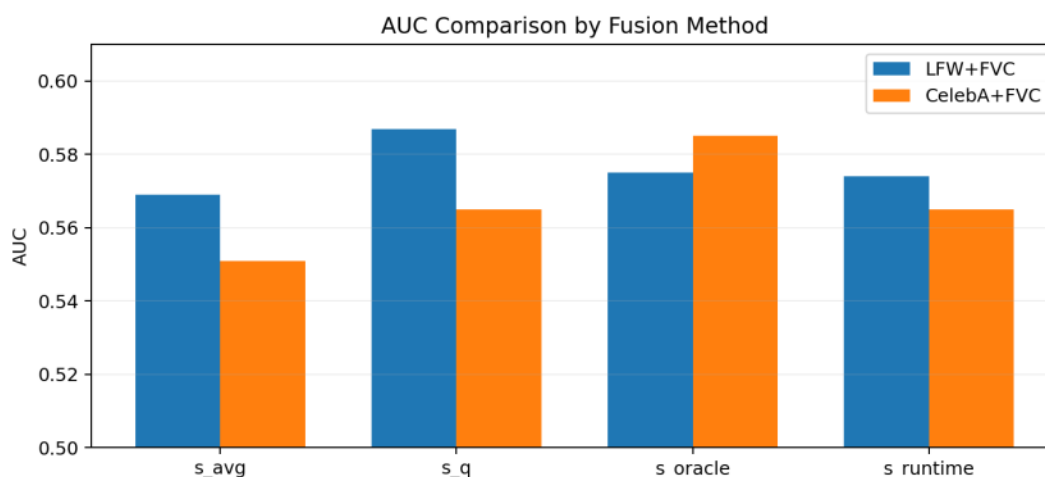


그림 6. 융합 방식별 인증 성능 비교

Fig. 6. Comparison of authentication performance across the fusion methods

+FVC의 s_oracle 은 TAR@FAR 5% 0.161로 가장 높았고, LFW+FVC에서는 s_oracle 과 $s_runtime$ 이 0.071로 공동 최고였다. 반면 ECE는 $s_runtime$ 이 CelebA+FVC 0.146, LFW+FVC 0.161로 각각 가장 낮았다.

다만 시험용 다른 대상 비교쌍이 56개뿐이므로 false positive 1개가 약 1.79% FAR에 해당한다. 따라서 TAR@FAR 1%와 0.1%는 표에 제시하되 통계적으로 안정적인 운영 성능으로 해석하지 않는다.

표 11의 AUC 95% 신뢰구간은 대부분 넓고 서로 중첩된다. 따라서 현재 표본만으로 특정 융합 방식의 통계적 우월성을 강하게 주장하기 어렵다. 이는 AUC 절대값이 낮다는 문제와 별개로 방식 간 차이의 검정력도 제한적임을 의미한다. 그림 6은 표 9의 AUC를 조합별로 비교한 것이다.

VI. 논의 및 한계

1. 품질과 식별 유용성 분리의 의미

본 연구에서 Q와 U의 선형 상관관계는 대부분 절댓값 0.07 이하로 낮았다. 이는 외형적 품질과 정답 방향 판정 여유가 동일한 척도가 아님을 시사하지만, 두 변수가 통계적으로 독립임을 증명한 결과는 아니다. 따라서 본 논문은 “독립

적”이라는 강한 표현 대신 “낮은 선형 상관” 또는 “상보적 정보를 포함할 가능성”으로 해석한다.

품질 기반 융합은 저품질 입력의 영향을 줄이는 데 유용하지만, 저품질이면서도 식별 정보가 남아 있는 경우를 충분히 반영하지 못할 수 있다. 반대로 고품질 입력이 항상 높은 판별력을 보장하지도 않는다. 이러한 관점에서 Q와 판정 기여 관련 정보를 분리해 표현하는 구조 자체는 후속 고성능 매체 환경에서도 검증할 가치가 있다.

2. 사후 분석용 oracle과 운영용 runtime의 구분

s_oracle 은 U_oracle 계산에 genuine/impostor 정답 정보를 사용하므로 실제 인증 시스템에 적용할 수 없다. 따라서 s_oracle 의 결과는 분석적 상한 또는 진단 결과로만 해석한다. 또한 LFW+FVC에서 s_oracle 이 s_q 보다 낮은 AUC를 보였으므로 oracle이라는 명칭이 모든 평가 지표의 최댓값을 보장한다는 의미도 아니다.

실제 운용 가능성을 판단하는 핵심은 $s_runtime$ 이다. 현재 결과에서 $s_runtime$ 은 AUC와 EER 측면에서 단순 평균 또는 품질 기반 방식에 대한 일관된 개선을 보이지 않았다. 다만 ECE에서는 두 조합 모두 가장 낮은 값을 나타내어, 현 단계에서는 “인증 정확도 향상 기법”보다는 “정답 없이 계산 가능한 신뢰도 보정 후보”로 해석하는 것이 타당하다.

표 12. 본 실험의 주요 한계와 보완 방향

Table 12. Main limitations of this study and directions for improvement

Limitation	Current Condition	Impact	Direction for Improvement
Sample size	112 test pairs, 56 impostor pairs	Unstable evaluation in the low-FAR regime	Extension with FVC and SOCOFing datasets
Face matching method	64×64 pixel similarity	Limited face discriminability	Adoption of ArcFace / FaceNet
Fingerprint matching method	ORB with Hamming distance	Insufficient representation of fine ridge structures	Adoption of minutiae-based matching or fingerprint embeddings
Multimodal configuration	Score-level simulation	Not based on real user pairs	Acquisition of face - fingerprint data from the same users
Statistical testing	Non-significant McNemar test	Insufficient statistical power	Larger samples and AUC-based comparison

3. 실험 한계와 보완 방향

본 연구는 제안 절차의 타당성을 확인하는 1차 검증으로 다음과 같은 한계를 갖는다. 첫째, 융합 시험은 112쌍, impostor 56쌍으로 제한되어 낮은 FAR 영역의 평가가 불안정하다. 둘째, 얼굴은 64×64픽셀 유사도, 지문은 ORB-Hamming 기반으로 최신 생체인식 매처에 비해 판별력이 제한된다. 셋째, 얼굴과 지문이 동일 사용자로 수집된 데이터가 아니므로 점수 기반 chimeric simulation이며 실제 사용자 단위 multimodal authentication의 상관구조를 반영하지 못한다. 넷째, AUC 신뢰구간이 넓고 서로 중첩되어 방식 간 차이의 통계적 확정성이 부족하다.

4. 향후 연구 방향

향후에는 얼굴 매처를 ArcFace, FaceNet, MagFace 등 사전 학습 임베딩 모델로 교체하고, 지문 매처는 minutiae 기반 매처 또는 지문 전용 딥러닝 임베딩으로 고도화할 필요가 있다. 또한 FVC의 여러 DB와 SOCOFing 등으로 지문 표본을 확장하여 수천~수만 쌍 규모의 시험을 구성해야 낮은 FAR 영역과 AUC 차이를 안정적으로 검증할 수 있다. 실제 운용을 위해서는 정답 정보를 사용하지 않고 식별 유용성을 추정하는 운영용 유용성 예측 모델이 필요하다. 품질점수, 판정 기준값과의 거리, 반복 촬영 안정성, 모달리티별 불확실성 등을 입력으로 학습하고 ECE, Brier score, reliability diagram을 이용해 보정하는 방향이 적절하다. 또한 동일 사용자에게서 얼굴과 지문을 함께 수집한 데이터셋으로 재검증하여 현재의 점수 수준 모의실험 한계를 해소해야 한다.

Ⅶ. 결 론

본 연구는 얼굴·지문 복합생체인증에서 입력 품질 Q와 판정 기여 관련 정보 U를 분리하고, 이를 신뢰도 기반 융합 가중치에 반영하는 방법을 제안하였다. Q, U, R과 모달리티 가중치의 정의식을 명시하고, 정답 정보를 사용하는 U_{oracle}/s_{oracle} 과 실제 운용 가능한 $C_{runtime}/s_{runtime}$ 을 구분하였다. 실험에서 Q와 U의 선형 상관관계는 낮았으나, U_{oracle} 이 정답을 이용한다는 점 때문에 판정 성공 여부와의 높은 관련성을 실용적 성능 증거로 해석하지 않았다.

CelebA+FVC에서는 s_{oracle} 이 AUC 0.585로 가장 높은 분석용 값을 보였지만, LFW+FVC에서는 s_q 의 AUC 0.587이 s_{oracle} 의 0.575보다 높았다. 운영용 $s_{runtime}$ 도 AUC/EER에서 일관된 우위를 보이지 않았으며, 두 조합에서 가장 낮은 ECE를 보인 점이 주요 관찰이다. 따라서 특정 융합 방식의 보편적 우수성을 주장하기보다, 품질과 판정 기여도를 분리해 해석하는 신뢰도 구조가 제한된 조건에서 보완 정보를 제공할 수 있다는 수준으로 결론을 제한한다.

현재 AUC가 0.5~0.6에 머문 것은 64×64픽셀 유사도와 ORB 기반 지문 매처, 시험 112쌍의 작은 표본, 서로 다른 얼굴·지문 데이터셋을 결합한 score-level simulation이라는 한계와 함께 해석해야 한다. 향후 고성능 얼굴·지문 매처, 대규모 시험쌍, 동일 사용자 기반 복합생체 데이터셋, 정답 없이 유용성을 예측하는 운영 모델을 적용한 재검증이 필요하다.

참 고 문 헌 (References)

- [1] P. L. Correia, A. Hadid, and S. Marcel, "Guest Editorial: Face Recognition and Spoofing Attacks," IET Biometrics, Vol.7, No.1, pp.1-2, January 2018.
doi: <https://doi.org/10.1049/iet-bmt.2017.0267>
- [2] A. F. Sequeira, M. Gomez-Barrero, N. Damer, and P. L. Correia, "BIOSIG 2021 Special Issue on Efficient, Reliable, and Privacy-Friendly Biometrics," IET Biometrics, Vol.11, No.5, pp.355-358, September 2022.
doi: <https://doi.org/10.1049/bme2.12101>
- [3] H. Es-Sobhah, M. Radouane, and K. Nafil, "Multimodal Biometrics: A Review of Handcrafted and AI-Based Fusion Approaches," IET Biometrics, Vol.2025, Article ID 5055434, October 2025.
doi: <https://doi.org/10.1049/bme2/5055434>
- [4] G. Chen, D. Luo, F. Lian, F. Tian, X. Yang, and W. Kang, "A Multimodal Biometric Recognition Method Based on Federated Learning," IET Biometrics, Vol.2024, Article ID 5873909, November 2024.
doi: <https://doi.org/10.1049/2024/5873909>
- [5] N. Poh and S. Bengio, "Can Chimeric Persons Be Used in Multimodal Biometric Authentication Experiments?," Proceeding of International Workshop on Machine Learning for Multimodal Interaction (MLMI), Lecture Notes in Computer Science, Vol.3869, pp.87-100, 2006.
doi: https://doi.org/10.1007/11677482_8
- [6] S. Bengio, C. Marcel, S. Marcel, and J. Mariéthoz, "Confidence Measures for Multimodal Identity Verification," Information Fusion, Vol.3, No.4, pp.267-276, December 2002.
doi: [https://doi.org/10.1016/S1566-2535\(02\)00089-1](https://doi.org/10.1016/S1566-2535(02)00089-1)
- [7] S. Bharadwaj, M. Vatsa, and R. Singh, "Biometric Quality: A Review of Fingerprint, Iris, and Face," EURASIP Journal on Image and Video Processing, Vol.2014, Article 34, pp.1-28, 2014.
doi: <https://doi.org/10.1186/1687-5281-2014-34>
- [8] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.815-823.
doi: <https://doi.org/10.1109/CVPR.2015.7298682>
- [9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.4690-4699, June 2019.
doi: <https://doi.org/10.1109/CVPR.2019.00482>
- [10] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A Universal Representation for Face Recognition and Quality Assessment," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp.14225-14234.
doi: <https://doi.org/10.1109/CVPR46437.2021.01400>
- [11] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep Learning Face Attributes in the Wild," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp.3730-3738, December 2015.
doi: <https://doi.org/10.1109/ICCV.2015.425>

저 자 소 개



이 지 상

- 2021년 5월 ~ 2023년 1월 : 중소기업기술정보진흥원 주임연구원
- 2023년 1월 ~ 현재 : 서울경제진흥원 선임
- 2023년 3월 ~ 현재 : 서울과학기술대학교 정보통신미디어공학과 박사과정
- ORCID : <https://orcid.org/0009-0004-8459-6039>
- 주관심분야 : 컴퓨터비전



박 구 만

- 1984년 2월 : 한국항공대학교 전자공학과
- 1986년 2월 : 연세대학교 전자공학과 공학석사
- 1991년 2월 : 연세대학교 전자공학과 공학박사
- 1991년 3월 ~ 1996년 9월 : 삼성전자 신호처리연구소 선임연구원
- 2006년 1월 ~ 2007년 8월 : Georgia Institute of Technology Dept. of Electrical and Computer Engineering, Visiting Scholar
- 2016년 1월 ~ 2017년 12월 : 서울과학기술대학교 나노IT디자인융합대학원
- 1999년 8월 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 교수
- ORCID : <https://orcid.org/0000-0002-7055-5568>
- 주관심분야 : 컴퓨터비전, 지능형실감미디어