

# 스마트폰 환경의 인증 성능 최적화를 위한 다중 생체인식 융합 기법 연구

문현준\*, 이민형\*\*, 정강훈\*  
세종대학교 컴퓨터공학부\*, 세종대학교 기계항공우주공학부\*\*

## Authentication Performance Optimization for Smart-phone based Multimodal Biometrics

Hyeon-Joon Moon\*, Min-Hyung Lee\*\*, Kang-Hun Jeong\*  
School of Computer Science and Engineering, Sejong University\*  
School of Mechanical and Aerospace Engineering, Sejong University\*\*

**요약** 본 논문에서는 스마트폰 환경의 얼굴 검출, 인식 및 화자 인증 기반 다중생체인식 개인인증 시스템을 제안한다. 제안된 시스템은 Modified Census Transform과 gabor filter 및  $k$ -means 클러스터 분석 알고리즘을 통해 얼굴의 주요 특징을 추출하여 얼굴인식을 위한 데이터 전처리를 수행한다. 이후 Linear Discriminant Analysis 기반 본인 인증을 수행하고(얼굴인식), Mel Frequency Cepstral Coefficient 기반 실시간성 검증(화자인증)을 수행한다. 화자인증에 사용하는 음성 정보는 실시간으로 변화하므로 본 논문에서는 Dynamic Time Warping을 통해 이를 해결한다. 제안된 다중생체인식 시스템은 얼굴 및 음성 특징 정보를 융합 및 스마트폰 환경에 최적화하여 실시간 얼굴검출, 인식과 화자인증 과정을 수행하며 단일 생체인식에 비해 약간 낮은 95.1%의 인식률을 보이지만 1.8%의 False Acceptance Ratio를 통해 객관적인 실시간 생체인식 성능을 입증하여 보다 신뢰할 수 있는 시스템을 완성한다.

**주제어** : 얼굴 검출, 얼굴 인증, 화자 인증, 동적 시간 정합, 다중 생체인식

**Abstract** In this paper, we have proposed personal multimodal biometric authentication system based on face detection, recognition and speaker verification for smart-phone environment. Proposed system detect the face with Modified Census Transform algorithm then find the eye position in the face by using gabor filter and  $k$ -means algorithm. Perform preprocessing on the detected face and eye position, then we recognize with Linear Discriminant Analysis algorithm. Afterward in speaker verification process, we extract the feature from the end point of the speech data and Mel Frequency Cepstral Coefficient. We verified the speaker through Dynamic Time Warping algorithm because the speech feature changes in real-time. The proposed multimodal biometric system is to fuse the face and speech feature (to optimize the internal operation by integer representation) for smart-phone based real-time face detection, recognition and speaker verification. As mentioned the multimodal biometric system could form the reliable system by estimating the reasonable performance.

**Key Words** : Modified Census Transform, Face Authentication, Mel Frequency Cepstral Coefficient, Dynamic Time Warping, Multimodal Biometrics

\* Authors are greatly acknowledging the support by Defense Acquisition Program Administration and Agency for Defense Development (UD120053GD).

Received 6 April 2015, Revised 10 May 2015

Accepted 20 June 2015

Corresponding Author: Kanghun Jeong (Sejong University)

Email: hereim80@gmail.com

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

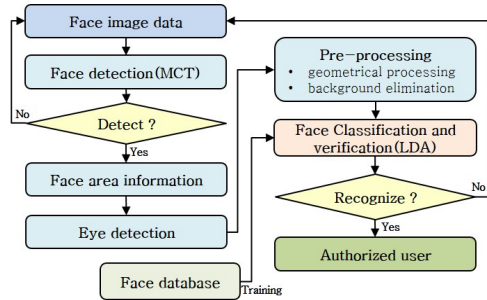
### 1. 서론

최근 스마트폰은 기술의 발전과 더불어 실생활에 밀접하게 연관되어 다양한 편의성을 제공하는 반면, 상당수의 개인정보를 저장하고 있다. 이에 따라, 일반적으로 사용하는 패스워드의 망각, 분실 및 복제 등에 취약한 단점을 해결할 수 있는 유일한 생체 데이터를 사용하는 생체인식 기반의 개인정보보호를 위한 방법에 대한 요구가 증가하고 있다. 기존 모바일 기기의 하드웨어 환경은 생체 데이터 처리에 필요한 연산량을 제공하지 못하였으나 최근 출시되는 스마트폰은 다중코어와 다양한 센서까지 보유하고 있어 생체인식 기술을 적용하기에 용이하다[1]. 다만, 조명변화 및 주변잡음 등 외부 환경변화에 영향을 쉽게 받는 특성상 기존에 출시되어 있는 얼굴인식 기술과 같이 한 가지의 생체특징을 사용하는 경우 인증 결과를 신뢰하기 어렵다. 따라서 다양한 생체 데이터를 복합적으로 사용하여 환경변화에 대응할 수 있는 다중생체인식 기술이 주목받고 있으나[2] 스마트폰 환경을 위한 다중생체인식 시스템은 아직 선보인 것이 없는 상황이다. 따라서 본 논문에서는 스마트폰에서 획득한 얼굴 및 음성정보를 융합한 스마트폰 환경의 얼굴 및 화자인증 기반 다중 생체인식 시스템을 제안한다. 제안된 시스템은 기존 시스템 대비 ±3% 이내의 인식률 차이와 실시간 처리 가능성을 목표로 한다.

### 2. 얼굴인증 시스템

얼굴특징 정보를 기반으로 신원을 인식하는 얼굴인식 기술은 지문인식과 같이 접촉식 센서를 사용하여 생체 정보를 획득할 필요가 없이 카메라를 이용하여 비 접촉식으로 데이터를 획득할 수 있다는 장점이 있다[3]. 얼굴 데이터는 생체 특징 중 개성이 뚜렷하여 생체인식에 매우 용이한 데이터로서, 비 접촉식 방식의 특성 상 사용자의 거부감이 적다[4]. 스마트폰 환경에서 얼굴인식 시스템을 구현하는 경우 영상 획득을 위한 추가적인 장치가 필요하지 않으며 사용자의 사용 패턴에 맞게 다양한 응용 어플리케이션을 구현할 수 있는 장점이 있다[5]. 본 논문에서는 Modified Census Transform(MCT) 기반의 얼굴검출 및 gabor filter[6]를 통해 눈을 검출하며, 눈의 좌표를 기준으로 전처리(pre-processing)를 수행하여 얼굴

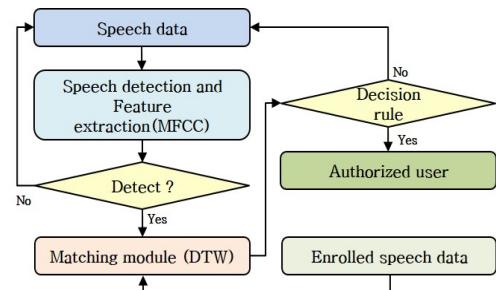
이외의 배경 데이터를 제거하고 정규화 된 이미지를 생성한다[7,8]. 등록된 사용자의 다중 얼굴데이터는 Linear Discriminant Analysis(LDA) 알고리즘을 통해 등록된 사용자의 얼굴만을 class화하여 1:1 matching function을 통해 인증한다[9]. 인증된 얼굴은 스마트폰 사용 환경을 고려한 실시간 추적을 통해 인증 정보를 유지한다.



[Fig. 1] Face verification system overview

### 3. 화자인증 시스템

얼굴인증 과정을 통과하면, 사용자가 등록해 둔 단어에 대한 음성신호를 통해 화자인증을 수행한다. 화자인증은 음성을 통해 본인을 판별하는 방법[10]으로써 전처리는 입력 음성을 frame단위로 나누어 energy 및 zero crossing rate(ZCR)을 분리하는(음성-묵음구간 분리) end-point detection(EPD)과 인간의 청각 특성을 반영한 Mel Frequency Cepstral Coefficient (MFCC)값을 추출하여 수행한다[11]. 전처리된 음성 데이터는 등록된 데이터와의 Dynamic Time Warping(DTW) 기반 유사도를 계산하여 임계값을 통해 본인 여부를 결정한다[12,13].

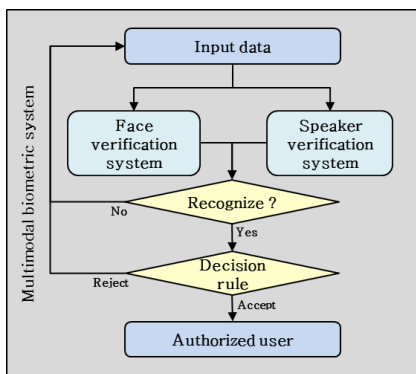


[Fig. 2] Speaker verification system overview

## 4. 생체인증 실험

### 4.1 실험환경 및 조건

다중 생체인식 실험은 Samsung Galaxy S (1GHz CPU, Android 2.3.6 platform)에서 20명의 얼굴 및 음성 정보를 기반으로 본인/타인인증 실험을 각 30회씩 수행하였다. 각 실험은 MCT기반 얼굴검출 및 LDA 기반의 얼굴인증 시스템을 통해 얼굴 유사도(distance)를 도출하고, 인증된 사용자에게 대해 입력된 음성 정보를 DTW기반 화자인증을 통해 입출력패턴의 유사도를 구한다[14]. 각 생체인식 실험은 170 ~ 220lx의 조도조건 및 45 ~ 50dB의 주변 소음 도를 보이는 실내 환경에서 획득한 생체 특징 벡터를 통해 유사도를 도출하였다. 유사도는 위의 알고리즘을 통해 추출한 생체 특징 벡터의 각 요소에 대해  $L_2 norm$  방식의 거리비교를 통해 계산한다.  $L_2 norm$ 은 생체인식 분야에서 인식 성공 여부를 객관적으로 평가할 수 있는 지표가 되는 대표적인 방법이다.



[Fig. 3] Multimodal biometric system overview

### 4.2 얼굴인증 실험

얼굴인증 실험을 위해 스마트폰 카메라로부터 각각 20명의 800×480 해상도의 영상을 획득하여 probe set(입력 얼굴 이미지 셋)로 사용하였으며 각 30회씩 실험하였다. Probe는 얼굴검출 및 전처리 수행 후, LDA 알고리즘을 통해 학습된 eigenspace(고유얼굴 공간)에 projection하여 가중치 벡터를 추출하여 gallery(등록된 사용자)의 가중치의 각 벡터 요소간의  $L_2 norm$  기반 유사도를 측정한다[15]. Probe와 gallery는 동일인의 경우에도 데이터 획득 일시가 다르며, 본인인증 실험 및 타인의 데이터를

이용한 무작위 인증의 유사도를 추출하였다. 실험 결과 본인인증은 평균 유사도 2794, 표준편차 74.5로 나타났으며, 타인인증은 평균 유사도 4963, 표준편차 180.5로써 본인 및 타인인증의 유사도의 차이로 인해 인증의 성공 여부를 구분 할 수 있는 명확한 결과를 보였다.

<Table 1> Face recognition result of enrolled user and the others

Probe set	A	B	C	D	E
Enrolled count	30	30	30	30	30
Average distance	2597.2	2891.1	2861.9	2817.5	2697.4
The others count	30	30	30	30	30
Average distance	4909.2	5213.8	5010.7	4527.2	5191.5

### 4.3 화자인증 실험

주어진 단어에 대해 30번씩 발음한 20명의 음성데이터를 8bit Mono 8000Hz Sampling format으로 저장하여 화자인증 실험을 진행하였다. 저장된 음성신호에서 MFCC기반 특징을 추출하여 저장된 음성신호와 DTW matching으로 유사도를 측정하여 각각 본인인증 실험 전체 평균(138.7)과 표준편차(5.7)를 구하였고 타인인증 실험에 대해 전체 평균(225.4), 표준편차(17.4)를 구하였다. 이 값을 토대로 명확히 본인 및 타인을 구분할 수 있음을 실험을 통해 확인하였다.

<Table 2> Speech recognition result of enrolled user and the others

Word	Seoul	University	gwanghwamun	Yi Sun-sin	choongmooogong
Enrolled count	30	30	30	30	30
Average distance	140.12	131.32	149.91	123.71	145.21
The others count	30	30	30	30	30
Average distance	259.30	211.19	247.13	217.48	264.16

## 5. 얼굴인증 시스템

### 5.1 고안된 다중생체인식의 결정 규칙

다음 수식 (1)을 적용하여 본인 및 타인인증 실험에서 얻은 평균  $\mu$ 와 표준편차  $\sigma$ 기반의 gaussian Probability Density Function(PDF)를 구한다. PDF는 본인( $\omega_1$ ) 및 타인( $\omega_2$ )인증의 결과를 가장 잘 표현할 수 있는 방법으로 써 [Fig. 4]는 얼굴인식의 본인인증 실험과 타인인증 실험

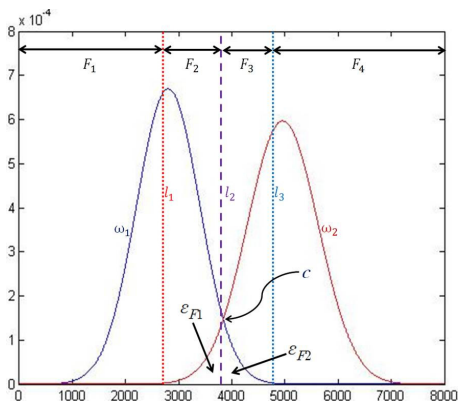
험결과의 PDF와 인증 성공 오류확률(False Acceptance Ratio; FAR,  $\epsilon_{F1}$ ) 및 인증 실패 오류확률(False Rejection Ratio; FRR,  $\epsilon_{F2}$ )을 각각 그래프의 영역으로 나타낸 것이다. Gaussian PDF와  $\epsilon_1$ 과  $\epsilon_2$ 는 다음 수식으로 구한다.

$$g_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

$$\epsilon_1 = \int_{-\infty}^{D_{L2}} p(x|\omega_2) dx, \quad l_1 \leq D_{L2} < l_2 \quad (2)$$

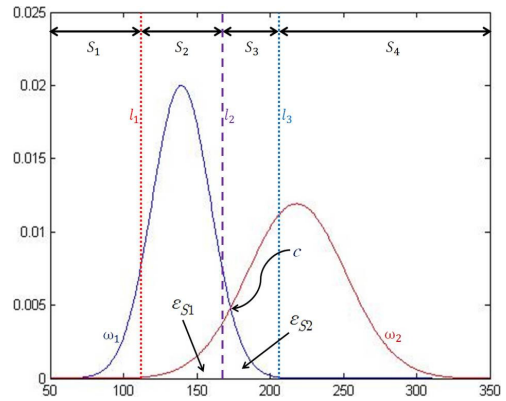
$$\epsilon_2 = \begin{cases} \left( 1 - \int_{-\infty}^c p(x|\omega_1) dx + \int_{-\infty}^c p(x|\omega_2) dx - \int_{-\infty}^{D_{L2}} p(x|\omega_2) dx \right) & \text{if } l_2 \leq D_{L2} < c \\ 1 - \int_{-\infty}^{D_{L2}} p(x|\omega_1) dx, & \text{else if } c \leq D_{L2} < l_3 \end{cases} \quad (3)$$

위의 식에서  $D_{L2}$ 는 각 생체인식 시스템의 유사도이며, 이를 이용하여 [Fig. 4]의  $\omega_1 \sim \omega_2$  영역을 선택 하여 최대 확률 값과 distance( $x$ 축)의 orthogonal line  $l_1, l_3$ 을 나타 낼 수 있다.  $l_2$ 는 최소 오류율(Equal Error Rate; EER)을 갖는 최적 결정 경계로써  $\epsilon_{F1}$ 과  $\epsilon_{F2}$ 의 넓이가 같은 지점이다. EER은 생체인식의 결과에 대해 승인 혹은 거부(accept or reject) 결정에 매우 중요한 요소이다.



[Fig. 4] Face recognition system result (left : enrolled user authentication, right : other authentication)

이와 같은  $l_1, l_2, l_3$ 를 기준으로 유사도 값을 각각  $F_1, F_2, F_3, F_4$  영역으로 나누고 PDF의 교차점을  $c$ 라고 할 때,  $l_1(2797.5), l_2(3832.9), l_3(4850.7), c(3940.7)$ 로 나타났다. [Fig. 5]은 얼굴인식 실험 결과와 같은 방법으로 화자인식의 PDF를 나타냈으며  $\epsilon_{S1}$  및  $\epsilon_{S2}$  영역의 넓이는 각각 화자인식에 대한 FAR과 FRR을 나타낸 것이다. 화자인식 실험 결과 각각  $l_1(109.70), l_2(168.06), l_3(208.92), c(173.06)$ 으로 나타났다.



[Fig. 5] Speaker recognition system result (left : enrolled user authentication, right : other authentication)

<Table 3> Proposed decision rule of biometric system

Face	Speaker	Decision rule, A : Accept, R: Reject
$F_1$	$S_{1-3}$	A
	$S_4$	R
$F_2$	$S_{1-2}$	A
	$S_3$	if $\epsilon_{F1} \geq \epsilon_{S2}$ R else A
$F_3$	$S_1$	A
	$S_{3-4}$	R
$F_4$	$S_2$	if $\epsilon_{F2} \geq \epsilon_{S1}$ R else A
	$S_{1-4}$	R

<Table 3>은 얼굴 및 화자인식의 4×4가지의 조합에 대해 본 논문에서 제시하는 다중생체인식 융합 최종 결정 규칙이다. 결정경계 기반 결정 규칙은 일부 생체인식 시스템이 비교적 낮은 인식률을 보이더라도 전반적인 인식성능을 보장할 수 있으며 인식 성능에 큰 영향을 주는

FAR을 감소시키기 위한 목적으로 입력 데이터에 대해 생체인식 시스템의 유사도 값이  $F_4$  혹은  $S_4$ 의 영역에 존재하는 경우 인증실패로 결정한다.

## 5.2 다중생체인식 실험 결과 및 분석

<Table 4>는 단일 및 다중 생체인식 시스템의 실험결과이다. 실험은 본인 및 타인인증 실험에 대해 20명을 대상으로 각 30회씩 수행하였다. 다중생체인식 시스템의 인식 결과는 얼굴 및 화자인식 시스템의 인식 결과가 각각 순차적으로 참일 경우에 인식 성공을 반환한다. 이를 직렬 다중생체인식 시스템이라 한다. 이와 같은 직렬 다중생체인식 시스템의 결정경계는  $I_2$ (EER)를 적용하여 판단하였으며, 얼굴인식의 FRR는 4.6%, 화자인식의 FRR은 6.7%로 나타났다. 제안한 다중생체인식 시스템의 FRR은 4.9%로써 주변 잡음에 매우 민감한 특성을 갖는 화자인식이 전체 FRR에 영향을 주어 얼굴인식만을 수행할 때의 FRR보다 0.3% 증가했지만, 화자인식만의 FRR에 비해서는 1.8% 감소했다.

이와 같은 이유는 제안된 다중생체인식 시스템에서 얼굴 및 화자인식의 순차 과정을 통해 최종 FRR을 계산하기 때문이다. 또한, 각 생체인식 시스템의 FAR을 비교하면 얼굴인식(4.6%), 화자인식(6.7%)으로 화자인식의 수치가 높게 나타나지만, 제안한 다중생체인식(1.8%)은 기존 방식에 비해 많이 감소한 것을 알 수 있다. 따라서 본 논문에서 제안하는 다중생체인식 시스템의 특성은 기존 생체인식 시스템에 비해 FRR수치는 비슷하며, FAR은 1.8%로 가장 작은 값을 보이는 신뢰할 수 있는 인증 시스템이다.

<Table 4> Comparison of multimodal and single recognition system

	FAR	FRR	인식률
Single face recognition	4.6%	4.6%	95.4%
Single speaker recognition	6.7%	6.7%	93.3%
Proposed multimodal biometrics	1.8%	4.9%	95.1%

## 6. 결론

본 논문에서는 기존 생체인식 시스템의 인식률 향상

및 실시간성 확인을 위해 얼굴 및 화자 인식 시스템 기반의 다중 생체인식 시스템을 구현하여 성능을 평가하였다. 제안된 시스템은 기존 시스템의 인식률(단일 얼굴인식(95.4%) 및 단일 화자인식(93.3%))에 비해 95.1%의 상대적으로 낮은 인식률을 보인다. 이 값은 단일 생체인식 기술 중 상대적으로 높은 인식률을 보이는 얼굴인식 기술에 비해 0.4% 낮은 인식률을 보이지만, 생체인식 시스템의 신뢰성과 직접적으로 관련이 있는 FAR값이 제안된 시스템의 경우 1.8%로써 기존 시스템 대비 신뢰할 수 있는 시스템임을 입증하였다. 낮은 FAR은 스마트폰 및 실외의 급격한 환경변화에 대응할 수 있음을 증명하는 근거로써 사람이 상주하지 않는 다양한 출입통제 시스템에 적용할 시 기존 시스템에 비해 높은 신뢰감을 줄 수 있다. 대표적인 예로는 실생활에서 쉽게 접할 수 있는 아파트 무인 현관부터 높은 수준의 보안이 요구되는 출입통제 구역에 이르기까지 다양한 환경에 적용 가능하다.

본 논문에서 제안한 다중생체인식 시스템은 얼굴 및 화자인증의 결정 규칙을 통하여 환경 변화에 강인한 얼굴검출 및 인식을 통해 본인인증을 수행하며, 화자인식 기반 실시간성 검증을 통해 단일 생체인식 시스템 대비 높은 신뢰성의 강인한 생체인식 성능을 보장한다.

## ACKNOWLEDGMENTS

Authors are greatly acknowledging the support by Defense Acquisition Program Administration and Agency for Defense Development (UD120053GD).

## REFERENCES

- [1] A. K. Jain, "Biometrics Personal Identification in Networked Society," Kluwer Academic Publishers, 1999.
- [2] Z. Akhtar, N. Alfarid, "Robustness of Serial and Parallel Biometric Fusion against Spoof Attacks", Communications in Computer and Information Science, Volume 157, pp 217-225, 2011.
- [3] A. Tolba, A. El-Baz, A. El-Harby, "Face Recognition: A Literature Review," Intern. Journ. of

Signal Processing, Vol. 2, No. 2, pp. 88-103, 2006.

[4] W. Zhao, "Face Recognition: A Literature Survey," ACM Computing Surveys, Vol. 35, No. 4, pp. 339-458, 2003.

[5] K. Jeong, S. Kim, and H. Moon, "Enhancement of Mobile Authentication System Performance based on Multimodal Biometrics," The 39th KIPS Spring Conference 2013.

[6] S. Xie, S. Shan, X. Chen, J. Chen, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition," IEEE Trans. on Image Processing, Vol. 19, No. 5, pp. 1349-1361, May 2010.

[7] Z. Qian, D. Xu, "Automatic Eye Detection Using Intensity Filtering and K-means Clustering," Pattern Recognition Letters, Volume 31, Issue 12, Pages 1633-1640, 1 September 2010.

[8] M. H. Yang, D. Kriegman, N. Ahuja, "Detecting faces in images: A survey," IEEE Transactions on Pattern Analysis and machine Intelligence, 24(1) : 34 - 58, January 2002.

[9] J. Ruiz-del-Solar, P. Navarrete, "Eigenspace- Based Face Recognition: A Comparative Study of Different Approaches," IEEE Trans. Systems, Man, and Cybernetics- Part C 35(3), 315-325, 2005.

[10] W. Shen, R. Khanna, "Prolog To Speaker Recognition: A Tutorial," Proceedings of the IEEE , vol.85, no.9, pp.1436, Sept. 1997.

[11] M. R. Hasan, M. Jamil, M. G. Rabbani, & M. S. Rahman, "Speaker identification using Mel frequency cepstral coefficients. variations," 3rd International Conference on Electrical & Computer Engineering ICECE 2004

[12] D. J. Berndt, J. Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series," In KDD workshop, Vol. 10, No. 16, pp. 359-370, 1994.

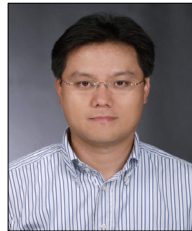
[13] D. A. Reynolds, "An overview of automatic speaker recognition technology," Proceedings of the International Conference on Acoustics, Speech, and Signal Processing(ICASSP)(S. 4072-4075). 2002

[14] E. Keogh, C. A. Ratanamahatana, "Exact indexing of dynamic time warping," Knowledge and

Information Systems, Volume 7, Issue 3, pp 358-386, March 2005.

[15] H. Moon, P. J. Phillips, "Computational and Performance Aspects of PCA-Based Face-Recognition Algorithms. Perception," Vol. 30. pp. 303-321, 2001.

**문 현 준(Moon, Hyeon Joon)**



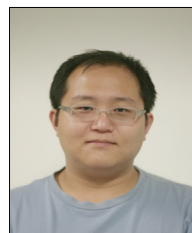
- 1990년 2월 : 고려대학교 전자공학과 (공학사)
- 1992년 6월 : SUNY Buffalo 컴퓨터공학과(석사)
- 1999년 9월 : SUNY Buffalo 컴퓨터공학과(박사)
- 1996년 2월 ~ 1999년 9월 : US Army Research Laboratory Researcher
- 1999년 2월 ~ 2003년 9월 : Viisage Technology Senior Researcher
- 2003년 2월 ~ 2004년 2월 : 연세대학교 조교수
- 2004년 3월 ~ 현재 : 세종대학교 전자정보공학대학 컴퓨터공학과 교수
- 관심분야 : 컴퓨터 비전, 패턴인식, 생체인식
- E-Mail : hmoon@sejong.ac.kr

**이 민 형(Lee, Min Hyung)**



- 1991년 2월 : 서울대학교 기계공학과(석사)
- 1995년 12월 : Univ. of Texas at Austin, 기계공학과 (박사)
- 1998년 3월 ~ 현재 : 세종대학교 기계항공우주공학부 교수
- 관심분야 : 전산역학, 비정상수치해석, 충돌/폭발현상
- E-Mail : mlee@sejong.ac.kr

**정 강 훈(Jeong, Kang Hun)**



- 2007년 2월 : 세종대학교 컴퓨터공학과 (공학사)
- 2009년 2월 : 세종대학교 컴퓨터공학과 (석사)
- 2009년 3월 ~ 현재 : 세종대학교 컴퓨터공학과 박사과정
- 관심분야 : 생체인식, 패턴인식, object tracking
- E-Mail : hereim80@gmail.com