

# 원통 모델과 스테레오 카메라를 이용한 포즈 변화에 강인한 얼굴인식

## (Pose-invariant Face Recognition using a Cylindrical Model and Stereo Camera)

노진우<sup>†</sup>    홍정화<sup>\*\*</sup>    고한석<sup>\*\*\*</sup>  
(Jin-Woo Noh) (Jung-Hwa Hong) (Hanseok Ko)

**요약** 본 논문에서는 원통모델과 스테레오 카메라를 이용하여 대상의 포즈 변화에 강인한 얼굴인식 방법을 제안한다. 입력으로 하나의 영상을 취할 수 있는 경우와 스테레오 영상을 취할 수 있는 경우의 두 가지로 나누어 다룬다. 단일 입력 영상인 경우 정면이 아닌 입력 영상에 대하여 원통 모델을 이용하여 좌우방향(yaw)으로 포즈를 보상하고, 스테레오 입력 영상인 경우 스테레오 기하학을 이용하여 예측된 상하방향(pitch) 포즈로 대상의 상하 변화까지 보상한다. 또한 스테레오 카메라를 통하여 동시에 두 개의 영상을 얻는다는 장점이 있기 때문에 결정 단계 융합(decision-level fusion) 방법을 이용하여 전체적인 인식률을 향상시킨다. 실험 결과, 좌우 포즈 변환을 통하여 인식률이 61.43%에서 94.76%로 향상되었음을 볼 수 있었고, 보다 복잡한 3차원 얼굴 모델과의 비교 결과 인식률이 양호함을 확인할 수 있었다. 또한 스테레오 카메라 시스템을 이용하여 얼굴이 위로 향한 영상일 경우 5.24%의 인식률을 향상시켰고, 결정 단계 융합에 의해 추가로 3.34%의 인식률을 향상시킬 수 있었다.

**키워드** : 얼굴 인식, 얼굴 포즈, 원통 모델, 스테레오 카메라

**Abstract** This paper proposes a pose-invariant face recognition method using cylindrical model and stereo camera. We divided this paper into two parts. One is single input image case, the other is stereo input image case. In single input image case, we normalized a face's yaw pose using cylindrical model, and in stereo input image case, we normalized a face's pitch pose using cylindrical model with previously estimated pitch pose angle by the stereo geometry. Also, since we have an advantage that we can utilize two images acquired at the same time, we can increase overall recognition performance by decision-level fusion. Through representative experiments, we achieved an increased recognition rate from 61.43% to 94.76% by the yaw pose transform, and the recognition rate with the proposed method achieves as good as that of the more complicated 3D face model. Also, by using stereo camera system we achieved an increased recognition rate 5.24% more for the case of upper face pose, and 3.34% more by decision-level fusion.

**Key words** : face recognition, face pose, cylindrical model, stereo camera

### 1. 서론

얼굴 인식은 지문 인식, 홍채 인식 등과 더불어 관심 받는 생체 인식의 한 분야이며, 공항 등 공공장소에서의 범죄자 검색, 출입 통제 시스템, ATM, 장난감 등 많은

분야에 이용될 수 있다. 얼굴 인식은 다른 인식 방법과 비교하였을 때 그 대상에게 특별한 동작을 요구하지 않는다는 장점을 갖고 있다. 하지만 그로 인해 동일한 대상에 대해 항상 같은 영상을 얻을 수 없는 문제가 발생한다. 따라서 이러한 변화에 강인한 인식을 수행하기 위해서는 그에 대한 보상이 필요하다. 널리 알려졌듯이 입력으로 들어오는 얼굴 영상에 가장 영향을 많이 주는 요소는 조명과 더불어 대상의 포즈이다. 이러한 대상의 포즈 변화에 대응하고자 시각에 기반한 접근 방법(view-based approach)과[1-3] 얼굴 포즈를 정규화(normalization) 하는 방법[1,4,5] 등이 제안되어왔다. 시

<sup>†</sup> 비 회 원 : 고려대학교 전자컴퓨터공학과, 현재 삼성전자 연구원  
jwnoh@ispl.korea.ac.kr

<sup>\*\*</sup> 비 회 원 : 고려대학교 제어계측공학과 교수  
hongjh32@korea.ac.kr

<sup>\*\*\*</sup> 정 회 원 : 고려대학교 전자컴퓨터공학과 교수  
hsk0@korea.ac.kr

논문접수 : 2003년 7월 2일

심사완료 : 2004년 4월 16일

각에 기반한 접근 방법은 일정한 포즈 각도마다 갤러리 영상(gallery image)이나 고유 공간(eigenspace)이 존재하여, 새로운 입력 영상이 들어왔을 때 그 포즈에 해당하는 갤러리 영상과 비교하거나, 해당하는 고유 공간을 이용하는 방법이다. 하지만 이러한 방법은 요구되는 갤러리 영상의 수가 많고, 각 포즈에 해당하는 고유 공간을 구성해야 하기 때문에 훈련에 요구되는 샘플의 수도 많고 그만큼 훈련 부하도 커진다. 얼굴 포즈를 정규화하는 방법은 입력 얼굴 영상을 정면 영상으로 바꾸어 인식하는 방법이다. 그러한 방법으로 어파인 변환(affine transform)이나[1,5] 3차원 얼굴모델(3D face model)을 이용하는 방법이 있다[4]. 어파인 변환은 대상의 깊이 정보를 고려하지 못하기 때문에, 위치 이동(translation), 스케일링(scaling), 평면상의 회전(in-plane rotation) 등의 평면상의 변환만이 가능하다. 따라서 깊이상의 회전 변화(rotation in-depth)를 갖는 얼굴 포즈를 보상하기에는 부적절하다. 3차원 얼굴 모델을 이용하여 보상하는 경우도 실제 정면 얼굴에 가까운 영상을 얻을 수 있는 하지만, 3차원 얼굴 모델의 변형(deformation) 과정에서 계산량의 부하가 크며 전체적으로 그 과정이 복잡하다. 따라서 깊이 정보를 고려한 3차원 변환의 효과를 가지면서 비교적 계산량이 적은 방법이 요구된다.

본 논문에서는 3차원 모델로 원통 모델을 사용하여 비교적 간단한 방법으로 정면에 가까운 영상을 획득하여 인식에 이용하고자 한다. 원통 모델은 실제 얼굴 모양과 약간의 차이는 있으나 얼굴을 표현할 수 있는 모델로써 포즈를 변환시키기 위한 회전을 생각했을 때 원통이 가장 간단하면서도 적절한 형태의 모델이라 할 수 있다.

먼저 2장에서는 입력으로 단일 영상을 획득하는 경우 포즈의 좌우 변환(yaw pose transform)만을 통하여 포즈를 정규화 하는 방법을 제안하고, 3장에서는 스테레오 영상을 획득하는 경우 포즈의 상하 변환(pitch pose transform)까지 적용하여 정규화 하는 방법 및 결정 단계 융합을 제안한다. 그리고 마지막으로 4장에서 결론을 맺는다.

**2. 원통 모델을 이용한 좌우 포즈 변환**

일반적으로 얼굴 인식에서 대상의 포즈를 다룰 때, 얼굴의 좌우 변화만을 고려하는 경우가 많다[2][3][6]. 따라서 이 장에서는 포즈의 변화를 좌우 변화만이 있는 것으로 제한하여 다루고자 한다.

**2.1 포즈 예측 모델**

좌우 포즈 변환을 위해 사용된 포즈 예측 모델은 그림 1에 나타내었듯이, 얼굴을 원통형이라 가정하고 얼굴의 좌우 폭과 두 눈의 위치를 이용한다. 본 논문에서는 이미 검출된 다양한 포즈의 얼굴 영상에서 더 나은 인

식물을 얻기 위한 방법에 초점을 맞추고 있으며, 포즈 예측에 필요한 얼굴 영역과 눈의 검출은 본 논문에서 다루고자 하는 주제에서 벗어나므로 이미 알고 있다고 가정한다. 대상의 포즈를 나타내는  $\theta_{yaw}$ 는 두 눈 사이의 중간점이 얼굴 좌우 폭의 중심으로부터 떨어진 거리  $a$ 와 얼굴 폭의 반을 나타내는  $r$ 로부터 구해질 수 있으며 그 식은 다음과 같다[3].

$$\theta_{yaw} = \arcsin(a/r) \tag{1}$$

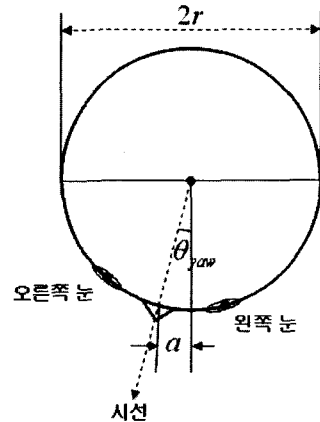


그림 1 포즈 예측 모델

**2.2 좌우 포즈 변환 과정**

대상의 포즈 예측 후, 얼굴을 원통형이라 가정하여 포즈 변환을 실시할 때 그 과정은 다음과 같으며, 구조도는 그림 3에 나타내었다. 먼저 입력 얼굴 영상의 각 픽셀 위치  $(x)$ 에 해당하는 깊이  $(z)$ 를 구한다. 원통 모델로 가정하였기 때문에,  $y$ 좌표에 따른 깊이 변화는 없으며 식 (2)로부터 식 (3)과 같은 깊이 정보를 얻을 수 있다.

$$(x-r)^2 + (r-z)^2 = r^2 \tag{2}$$

$$z = r - \sqrt{x(2r-x)} \tag{3}$$

이렇게 원통 모델을 사용하여 획득한 좌표  $(x, z)$ 를 예측된 포즈 각도  $\theta_{yaw}$ 만큼 회전시켜서 새로운 좌표  $(x')$ 를 구한다.

$$x' = (x-r)\cos\theta_{yaw} + (z-r)\sin\theta_{yaw} - r \tag{4}$$

그 후 입력 얼굴 영상의  $(x)$ 좌표에 해당하는 픽셀값을 생성된 좌표  $(x')$ 에 매핑 시키고, 변환으로 인해 생성된 빈 픽셀 값을 갖는 위치에 대해서는 양쪽 픽셀 값의 평균으로 보상하고, 정보가 없는 끝 부분의 영향을 제거하기 위해 양측을  $w$ 만큼 제거하면 변환이 완료된다. 포즈 변환 결과는 그림 3에 나타내었다. 포즈 변환 결과를 살펴보면, 원통 모델의 가정을 따라 영상이 비선형적으로 줄거나 늘어났음을 볼 수 있다.

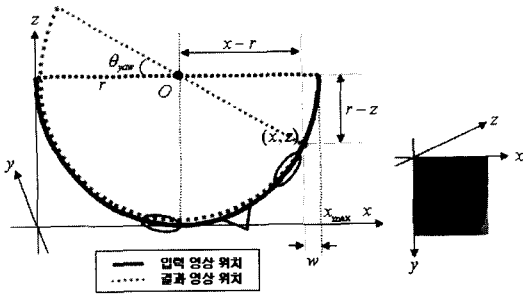


그림 2 좌우 포즈 변환 구조도

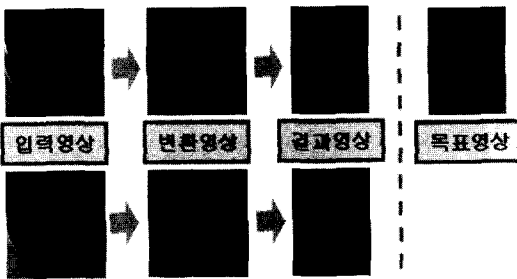


그림 3 좌우 포즈 변환 결과

2.3 실험 결과

2.3.1 좌우 포즈 변환으로 인한 인식을 향상 평가

좌우 포즈 변환으로 인한 인식을 향상을 평가하기 위해 주 성분 분석법(principal component analysis, PCA)를 사용하였다[7,8]. 주 성분 분석법이란 M차원상에 존재하는 영상을 분산을 최대로 하는 N개의 고유공간으로 사영하여 특징값을 얻는 방법으로써, 간단하면서도 효과적이기 때문에 얼굴인식에서 일반적으로 많이 사용된다. 이 논문의 경우 다양한 포즈에 대한 얼굴인식에 초점을 맞추고 있으므로, 원통모델을 이용한 포즈 변환방법의 성능을 시험해 보기 위해 가장 일반적으로 사용되는 PCA방법을 사용하였다. 이 논문에서는 주성분 분석법을 이용해 훈련영상의 특징값을 얻은 후 각 사람별 평균값을 계산해 놓고, 실험영상이 들어왔을 때 얻어진 특징값과 유클리디안 거리가 가장 가까운 거리에 있는 사람으로 인식하도록 했다. 각 실험에 사용된 대상은 21명이며, 훈련 영상으로는 대상 당 10개, 테스트 영상으로는 대상 당 10개를 사용하였고, 사용된 고유공간의 수는 50개이다.

표 1 좌우 포즈 변환으로 인한 인식을 변화(훈련 영상 각도 범위와 테스트 영상 각도 범위가 같을 경우)

범 위	-10°~+10°		-20°~+20°		-30°~+30°	
	변환전	변환후	변환전	변환후	변환전	변환후
인식률	87.62%	96.19%	67.62%	98.57%	61.43%	94.76%

표 1에서는 다루는 대상의 포즈 범위에 따른 포즈 변환 전과 변환 후의 인식률의 차이를 보여주고 있다(영상에서 -부호는 대상의 시선이 좌측, +부호는 우측인 경우를 가리킨다). -10°~+10° 범위의 포즈 변화를 다룰 경우 변환으로 인해 8.57%의 인식률 향상이 있음을 확인할 수 있으며, 다루는 대상의 포즈 범위가 넓어질수록 변환 후 인식률의 향상 폭이 더욱 커져 변환으로 인한 효과가 더욱 증가함을 볼 수 있다. 참고로 어떠한 포즈 범위를 다룬다는 것은, 그 포즈 범위 내의 영상들로 훈련하였다는 것을 의미한다.

2.3.2 3차원 얼굴 모델을 사용했을 때와의 비교 실험

3차원 얼굴 모델을 사용했을 때의 인식 결과와 비교하기 위하여, Jebara의 실험 결과와 비교하였다[4]. 데이터베이스로는 Achermann DB를 사용하였다. Achermann DB는 30명의 대상에 대해 각각 10개의 영상을 갖고 있으며, 정면(#1, #2), 오른쪽(#3,#4), 왼쪽(#5, #6), 위(#7, #8), 아래(#9, #10)의 영상으로 구성되어 있다(그림 4).

Jebara의 3차원 얼굴 모델을 이용한 인식 결과와 원통 모델을 이용한 인식 결과를 표 2에 나타내었다. 각 대상의 1번 영상 30개로 훈련하고, 전체 300개의 영상으로 테스트 한 결과이다. 원통 모델을 사용하였을 때 18.67%의 인식률 향상이 있음을 볼 수 있다. 그러나 Jebara의 실험에서는 실제 얼굴 검출 방법까지 적용하여 인식률을 얻었고, 원통 모델 실험에서는 얼굴과 눈의 위치를 정확히 알고 있다고 가정하고 실험을 하여 인식률을 얻었기 때문에 두 개의 인식 결과를 직접적으로 비교할 수 없다. M. H. Yang et al.[9]에 의한 조사에서 따르면 실험을 통해 보았을 때 일반적인 얼굴 검출률이 약 90% 이상의 성능을 보이므로[9] 이를 적용하여 비교한다. 원통 모델로 획득한 84.00%의 인식률에 일반적인 얼굴 검출률 90%를 적용하면 최종적인 인식률은 75.60%이므로, Jebara의 3차원 얼굴 모델보다 약 10%의 인식률 향상이 있게 된다. 3차원 얼굴 모델보다 간단한 모델인 원통 모델을 사용했을 때 인식률이 더욱 향

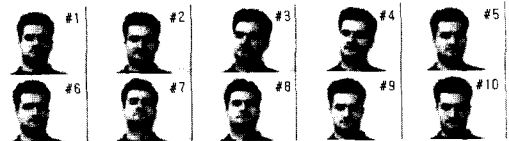


그림 4 Achermann DB의 대상의 포즈 예

표 2 Jebara의 3차원 얼굴 모델과 원통 모델을 사용했을 때 인식률 비교

	3차원 얼굴 모델	원통 모델
인식률	65.33%	84.00%

상된 이유는, Jebara가 계산량의 부하를 줄이기 위해 모델의 변형(deformation) 과정을 간략화 하여 수직 방향의 길이 조정만을 했기 때문인 것으로 보인다. 따라서 생성된 결과 영상은 실제 정면 영상과 같은 모습을 보이기는 하지만, 인식률에 영향을 미치는 대상 간의 변화(variation)가 줄어들어 인식률이 감소한 것으로 보인다. 그림 5의 (a)를 (b)와 비교해보면, 결과 영상 간의 변화가 줄어든 것을 볼 수 있고, 또한 결과 영상들의 평균 영상인 (c)가 선명한 것을 볼 때 결과 영상 간의 변화가 적은 것을 확인할 수 있다. 이와 반대로 원통 모델의 경우, 그림 3에서 볼 수 있듯이 생성된 결과 영상이 실제 정면과는 다른 왜곡된 모습을 보이기는 하지만 양호한 인식률을 갖는다. 이는 전체적으로는 왜곡되어 보이지만 각 부분적으로는 왜곡이 크지 않고, 또한 포즈변환을 통해 눈, 코, 입 등이 적절한 곳으로 위치하게 되기 때문으로, 비록 영상은 왜곡되어 보이지만 인식률은 향상된다.

그림 6은 대상의 포즈 변화에 따른 인식률을 보이고 있다. 포즈 번호는 그림 4에서 볼 수 있듯이 대상의 포즈를 가리킨다. 3차원 얼굴 모델을 사용할 경우 대상의 상하 포즈 변화(#7, #8, #9, #10)가 있을 때 인식률이 비교적 낮은 것을 볼 수 있다. 원통 모델을 이용하여 포즈의 좌우 변환을 하는 경우에도, 상하 변화를 고려하지 않기 때문에 그러한 영상들에 대해서 인식률이 낮아지게 된다. 하지만 그림 6을 보면, 아래로 향하는 영상(#9, #10)에 대한 인식률이 높게 나왔음을 볼 수 있다. 그 이유는 Achermann 데이터베이스를 살펴보면 아래로 향하는 영상들이 정면 영상에 가깝게 촬영된 경향이 있기 때문인 것으로 보인다. 하지만 원통 모델과는 달리 3차원 얼굴 모델을 이용했을 때 #9, #10의 영상들에 대해서 인식률이 크게 떨어지는 이유는 Jebara의 논문만으로는 더 이상 분석하기 어렵다.

지금까지 원통 모델을 이용한 좌우 포즈 변환을 통하여, 대상의 좌우 포즈 변화에 강인한 인식 결과를 얻을 수 있음을 보았다. 하지만 그림 6에서 볼 수 있듯이 위

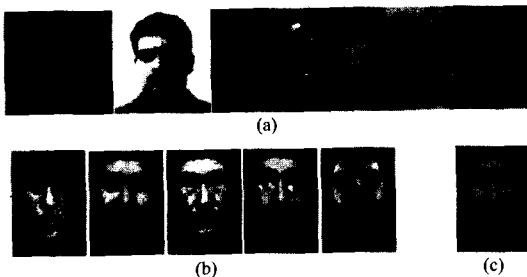


그림 5 Jebara의 3차원 얼굴 모델을 이용한 결과 (a)입력 영상의 예, (b)결과 영상의 예, (c)평균 영상의 예

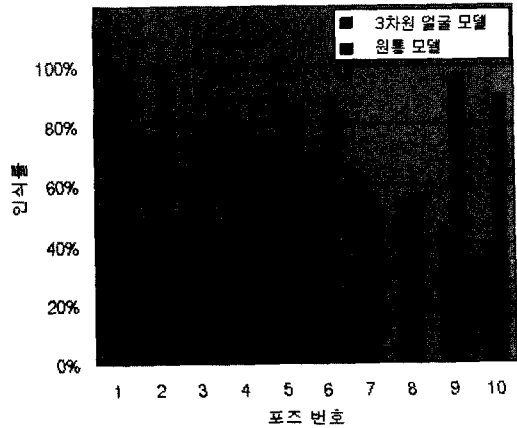


그림 6 포즈 변화에 따른 인식률 비교

로 향한 영상(#7, #8)의 경우 인식률이 떨어지고, 아래로 향한 영상의 경우에도 그림 6의 결과와는 달리 인식률이 떨어질 것이라 예상된다. 따라서 상하 포즈 변화에 대한 보상을 통하여 이러한 영상들에 대한 인식률을 향상시킬 필요가 있다. 이를 위하여 우리는 스테레오 카메라 시스템을 이용하고자 한다.

### 3. 스테레오 카메라를 이용한 상하 포즈 변환

얼굴의 상하 포즈 변화에 대한 보상을 하기 위해서는 우선 상하 포즈 예측이 필요하다. 단일 영상으로 얼굴의 상하 포즈를 예측하기 위해서 인체측정학(anthropometry)[10]이나 기준이 되는 정면 영상을 이용하는 방법[11]이 있다. 하지만 전자의 경우 인종 및 성별 등을 기준으로 얼굴의 구조를 분석하였기 때문에 인식 이전에 그 범주를 판단하는 과정이 필요하며, 또한 같은 범주로 분류되었다 하더라도 실제 우리가 접할 수 있는 얼굴의 다양성을 포함하기 어렵다. 후자의 경우도 인식하고자 하는 대상의 정면 영상을 필요로 하기 때문에 얼굴 인식에 적용하기에는 어려움이 있다. 이렇듯 단일 영상을 이용한 예측 방법들의 한계로 인해, 우리는 스테레오 카메라를 이용하여 상하 포즈를 예측하고자 한다 [12,13]. 또한 스테레오 카메라 시스템은 동시에 두 개의 영상을 얻을 수 있기 때문에, 결정 단계 융합(decision-level fusion)을 통하여 전체적으로 인식률을 향상시킬 수 있다는 장점도 있다.

#### 3.1 상하 포즈 예측 방법

스테레오 시스템을 이용하여 포즈를 예측하는 방법은 다음과 같다. 각 카메라의 캘리브레이션(calibration)을 통하여 내부 변수(intrinsic parameters)를 획득하고, 두 카메라와 동시에 연관된 임의의 월드 좌표계(world coordinate)를 기준으로 각 카메라에 해당하는 외부 변

수(extrinsic parameters)를 획득한다. 이러한 정보들과 양측 영상의 대응점(corresponding point)을 이용하여 양쪽 눈 끝, 입 끝의 4개 특징점의 3차원 좌표를 획득하고, 이렇게 획득된 3차원 좌표를 각 카메라 기준의 좌표로 변환 후, 특징점들이 구성하는 평면의 정규 벡터(normal vector)를 이용하여 포즈를 예측한다.

내부 변수와 외부 변수로 이루어지는 투사 행렬(projection matrix)  $P, P'$ 를 획득하기 위해서, OpenCV 라이브러리의 캘리브레이션 필터를 사용하였다.<sup>1)</sup> 또한 각 카메라의 투사 행렬 획득 후 양 측 영상의 대응점으로부터 그 특징점의 3차원 좌표를 복원하기 위하여 선형 삼각법(linear triangular method)[13]을 이용하였다 [13]. 각 영상으로부터 특징점의 좌표  $m = PM, m' = P'M$ 을 얻을 수 있고 각각의 식은 식 (5)와 같이 나타내어진다. 여기서  $m, m'$ 는 각각 좌측 및 우측 영상에서의 특징점의 2차원 좌표,  $M$ 은 복원할 특징점의 3차원 좌표이며, 동차(homogeneous) 좌표계로 표현하였다.

$$m = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = PM = \begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix} M \quad (5)$$

식 (5)로부터 아래의 3개의 식을 얻을 수 있다.

$$\begin{aligned} u(p_3^T M) - w(p_1^T M) &= 0 \\ v(p_3^T M) - w(p_2^T M) &= 0 \\ u(p_2^T M) - v(p_1^T M) &= 0 \end{aligned} \quad (6)$$

이렇게 양 측에서 얻을 수 있는 6개의 식을 이용하여 다음과 같은  $TM=0$  형태의 식을 얻을 수 있다.

$$\begin{bmatrix} u p_3^T - w p_1^T \\ v p_3^T - w p_2^T \\ u' p_3^T - w' p_1^T \\ v' p_3^T - w' p_2^T \end{bmatrix} M = TM = 0 \quad (7)$$

결국 특징점의 3차원 좌표  $M$ 은 행렬  $T$ 의 영벡터(null-vector)이며, 최소 제곱 근(least-squares solution)으로부터 구해질 수 있다. 대상의 포즈는, 이상의 방법으로 획득된 양 눈 끝과 입 끝의 4개의 특징점이 이루는 평면의 정규 벡터를 이용하여 구할 수 있으며, 그 과정은 다음과 같다. 우선 획득된 입의 월드 좌표계에서의 3차원 좌표를 각각의 카메라 좌표계에서의 좌표로 변환한다.

$$M_C = + t \quad (8)$$

$M_C$ 는 각 카메라 좌표계에서의 3차원 좌표를 나타낸다. 4개의 특징점의 좌표를 각각  $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3), (x_4, y_4, z_4)$ 라 하고, 평면의 방정식을

$ax + by + cz + d = 0$ 이라 하면 다음의 관계를 만들 수 있다.

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 0 \quad (9)$$

식 (9)로부터 최소 제곱근을 이용하여 정규 벡터를 구한다. 정규 벡터  $N$ 은

$$N = a' \hat{x} + b' \hat{y} + c' \hat{z} \quad (10)$$

$$\text{where } a' = \frac{-a}{d}, b' = \frac{-b}{d}, c' = \frac{-c}{d}$$

결국  $\theta_{pitch}, \theta_{yaw}$ 는 각각 정규 벡터와 X-O-Z 평면, Y-O-Z 평면이 이루는 각도와 같다.

$$\begin{aligned} \theta_{pitch} &= \arcsin \frac{b'}{\sqrt{b'^2 + c'^2}} \\ \theta_{yaw} &= \arcsin \frac{a'}{\sqrt{a'^2 + c'^2}} \end{aligned} \quad (11)$$

포즈 예측 모델은 그림 7에 나타내었다.

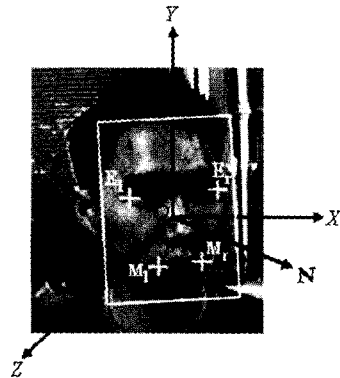


그림 7 특징점의 3차원 좌표를 이용한 포즈 예측 모델

### 3.2 상하 포즈 변환 과정

입력 영상의 상하 시선은 스테레오 시스템을 이용하여  $\theta_{pitch}$ 로 예측되었다. 따라서 입력 얼굴을  $\theta_{pitch}$ 만큼 회전된 원통형이라 가정하고, 그 모델에 맞춰 입력 영상의 각 픽셀 위치  $(x, y)$ 에 해당하는 깊이  $(z)$ 를 구한다 (그림 8).

$$\begin{aligned} z' &= (M - y) \sin \theta_{pitch} + z \cos \theta_{pitch} \\ &= (M - y) \sin \theta_{pitch} + (-r - \sqrt{x(2r - x)}) \cos \theta_{pitch} \end{aligned} \quad (12)$$

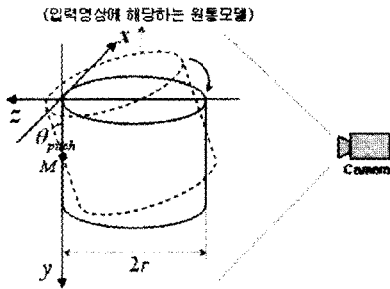
$r$ 은 원통의 반지름,  $z$ 는 정면인 원통 모델에서의 깊이,  $M$ 은 원통 회전의 기준점을 나타낸다. 이렇게 구한 3차원 좌표  $(x, y, z')$ 를  $\theta_{pitch}$ 만큼 회전시켜서, 새로운 좌표  $(x, y', z')$ 를 구한다.

$$y' = -(M - y) \cos \theta_{pitch} + z' \sin \theta_{pitch} + M \quad (13)$$

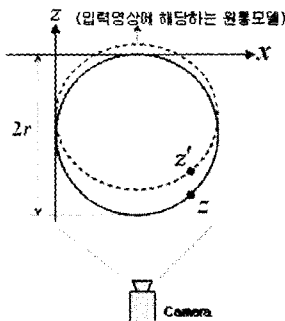
물론 원통 모델의 상하 회전에 따라 원근에 의해서

1) Intel Corp.에서 제공하는 OpenCV 라이브러리는 <http://www.intel.com/research/mrl/research/opencv/>에서 구할 수 있다.

원통 폭의 변화가 발생하기는 하지만 아주 미비하기 때문에 상하 변화로 인한  $x$  좌표의 변화는 없는 것으로 가정하였다. 그 후 입력 얼굴 영상의  $(x, y)$ 좌표에 해당하는 픽셀 값을 변환된  $(x', y')$ 에 매핑시키고, 회전으로 인해 생성된 빈 픽셀 값은 양측 픽셀의 평균값으로 보상하고, 생성된 영상에서 정보가 없는 상하측 부분을 제거한다. 상하 포즈 변환 결과는 그림 9에 나타내었다. 원통 모델의 가정을 따라 대략적으로 정면에 가깝게 변환된 모습을 볼 수 있지만, 얼굴을 바라보는 각도에 따



(a)



(b)

그림 8 원통 모델을 이용한 상하 포즈 변환 (a)대상을 측면에서 바라볼 때, (b)대상을 위에서 바라볼 때

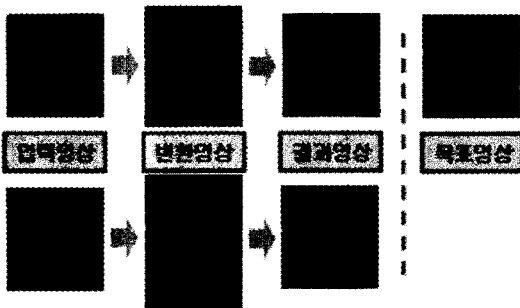


그림 9 상하포즈 변환 결과

라 실제 눈, 코의 변화가 크기 때문에 변환에 어려움이 있는 것으로 보인다.

### 3.3 결정 단계 융합(decision-level fusion)

스테레오 카메라를 사용하면 동일한 대상에 대해서 다른 포즈의 영상을 동시에 획득할 수 있기 때문에, 결정 단계 융합을 통하여 인식을 향상시킬 수 있다. 즉 왼쪽, 오른쪽 입력 영상 각각의 인식 결과가 다를 때 융합을 통하여 전체적인 인식률을 높이는 것이다. 스테레오 카메라에 결정 단계 융합을 적용한 구조를 그림 10에 나타내었다. 양 측 카메라로부터의 각각의 인식 결과는 융합 과정을 통하여 최종적인 인식 결과를 도출하게 된다.

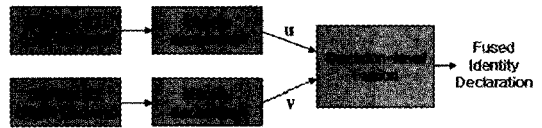


그림 10 결정 단계 융합 구조도

융합 과정의 입력으로 들어오는 각 센서(sensor)로부터의 인식 결과는 소프트 디시전(soft decision)으로서, 각 신원(identity)에 관한 확률값을 부여한 다중 선언(multiple declaration)을 사용하였다. 융합 과정의 입력  $u, v$ 는 각각  $M \times 1$ 의 벡터로서( $M$ 은 대상의 수), 입력 영상에서 추출된 PCA 특징과 데이터베이스 내의 각 대상의 PCA 특징과의 유클리드 거리(euclidean distance)가 작은 순서대로, 그 대상을 나타내는 번호를  $u_1$ 부터  $u_M$ 에(또는  $v_1$ 부터  $v_M$ 에) 할당한다.

$$\begin{aligned} u &\in N^{M \times 1}, \quad u = [u_1 \ u_2 \ \dots \ u_M]^T \\ v &\in N^{M \times 1}, \quad v = [v_1 \ v_2 \ \dots \ v_M]^T \end{aligned} \quad (14)$$

여기서  $N$ 은 자연수를 의미한다. 결정 단계 융합을 위하여, 실험적으로 각 인식 순위에 해당하는 확률값을 얻을 수 있으며 다음과 같이 표현된다.

$$\begin{aligned} P(u_1 = H_i | H_i) &= p_1 \\ P(u_2 = H_i | H_i) &= p_2 \\ P(u_3 = H_i | H_i) &= p_3 \\ &\vdots \\ P(u_M = H_i | H_i) &= p_M \end{aligned} \quad (15)$$

$$\sum_{i=1}^M P(u_i = H_i | H_i) = p_1 + \dots + p_M = 1 \quad (16)$$

$H_i$ 는 관측된 존재(entity)가 대상  $\#i$ 인 것을 나타낸다. 입력 영상이 주어졌을 때 인식단을 통과하게 되면 앞에서 기술한 바와 같이  $M \times 1$  크기의 벡터를 획득하게 되고, 그 벡터의 각 요소는 데이터베이스 내의 각 대상에 대한 인식 순위를 나타낸다. 그것으로부터 확률값  $p_i$  ( $i=1, \dots, M$ )은 다음과 같이 획득된다.

$$p_i = \frac{n_i}{T} \quad (17)$$

여기서  $T$ 는 테스트 영상의 수를 의미하며,  $n_i$ 은 입력으로 들어온 대상이 결과 벡터 내에서  $i$ 번째 요소인 경우의 횟수를 의미한다. 또한 융합을 통하여 얻고자 하는 것은 각 대상의 결합 확률(joint probability)  $P(H_i | \mathbf{u}, \mathbf{v})$ 이며, 그 중에서 가장 큰 값을 갖는 대상을 최종 인식 대상으로 결정하는 것이다.  $P(H_i | \mathbf{u}, \mathbf{v})$ 는 다음과 같이 표현될 수 있다.

$$P(H_i | \mathbf{u}, \mathbf{v}) = \frac{P(\mathbf{u}, \mathbf{v} | H_i) P(H_i)}{P(\mathbf{u}, \mathbf{v})} \quad (18)$$

여기서  $P(H_i)$ 는 사전 확률(priori probability)로서 모든 대상에 동일하게  $1/M$ 의 값을 갖으며,  $P(\mathbf{u}, \mathbf{v})$ 도 항상 일정한 값이기 때문에,  $P(H_i | \mathbf{u}, \mathbf{v})$ 을 다음과 같이 표현 가능하다.

$$\begin{aligned} P(H_i | \mathbf{u}, \mathbf{v}) &\propto P(\mathbf{u}, \mathbf{v} | H_i) \\ &= P(\mathbf{u} | H_i) \cdot P(\mathbf{v} | H_i) \\ &= P(u_j = H_i | H_i) \cdot P(v_k = H_i | H_i) \end{aligned} \quad (19)$$

$\mathbf{u}, \mathbf{v}$ 가 서로 독립적이기 때문에 첫 번째 행에서 두 번째 행으로 바뀔 수 있으며, 세 번째 행의  $j$ 와  $k$ 는 벡터  $\mathbf{u}, \mathbf{v}$ 에서 대상 #에 해당하는 위치를 의미한다. 결국 인식 결과의 확률값을 이용하여 대상 #를 찾는 과정은 다음과 같이 표현된다.

$$\arg \max_i P(H_i | \mathbf{u}, \mathbf{v}) \quad (20)$$

결정 단계 융합에서 마지막으로 확인해야 할 것은 두 명의 대상이 동일한 결합 확률을 갖는 경우이다. 이 경우 대상의 포즈가 정면에 가까울수록 신뢰성이 있다는 점을 이용하여 보다 정면에 가까운 영상을 입력으로 갖는 쪽의 결과를 따른다.

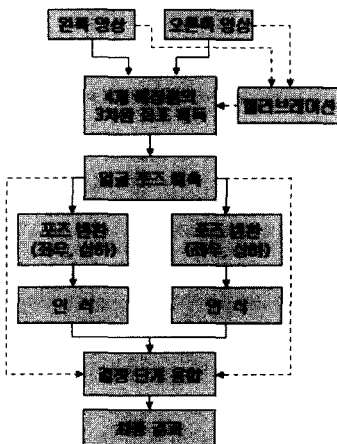


그림 11 제안한 얼굴 인식 과정

지금까지 제안한 원통 모델과 스테레오 카메라를 이용한 얼굴인식 과정을 요약하면 그림 11과 같다.

### 3.4 실험 결과

스테레오 시스템을 구축하기 위하여 CCD 기반의 웹 카메라(web camera) 2대를 사용하였다. 카메라의 해상도는  $320 \times 240$ , 시야각은  $62^\circ$ 이며, 스테레오 카메라 세팅 구조도는 그림 12와 같다. 각 카메라가 바라보는 각도  $\theta$ 는 약 50cm 떨어진 대상의 행동 제한을 최소화 할 수 있는 각도로 설정하였다.

$$\theta = \tan^{-1}\left(\frac{L/2}{D}\right) = \tan^{-1}\left(\frac{10}{50}\right) \approx 11.3^\circ \quad (21)$$

각 카메라의 캘리브레이션을 위하여 OpenCV 라이브러리를 사용하였으며, 캘리브레이션 물체(calibration object)로는 그림 13과 같은 체스판 패턴을 사용하였다. 내부 변수는 각 카메라 당 캘리브레이션을 20회 실시하여 평균을 취하였으며, 각 카메라에 해당하는 내부 변수를 포함한 행렬로써 다음과 같은 값을 획득하였다.

$$A = \begin{bmatrix} 362 & 0 & 171 \\ 0 & 361 & 95 \\ 0 & 0 & 1 \end{bmatrix} \quad A' = \begin{bmatrix} 364 & 0 & 184 \\ 0 & 362 & 135 \\ 0 & 0 & 1 \end{bmatrix} \quad (22)$$

임의의 월드 좌표계로부터 각 카메라 좌표계에 이르는 회전 행렬(rotation matrix)  $R, R'$ 와 이동 벡터(translation vector)  $t, t'$ 는 외부 변수로서, 두 카메라를 이용

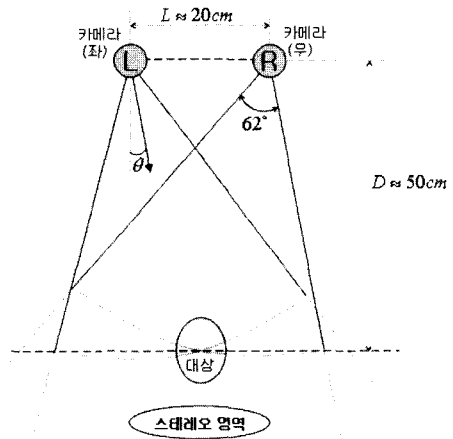


그림 12 스테레오 카메라 세팅

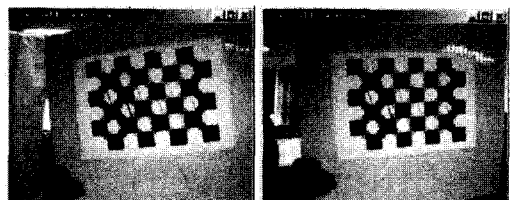


그림 13 사용된 캘리브레이션 물체

하여 동시에 획득하였을 시 카메라 간의 관계를 포함하게 된다. 따라서 캘리브레이션 물체를 이용하여 임의의 위치로부터 획득되는 값을 동시에 얻어 사용하였다.

상하 포즈 변환을 적용한 인식률을 표 3에 나타내었다. 실험에 이용된 대상의 수는 21명이며, 훈련 영상은 대상 당 평행 시선의  $-30^{\circ} \sim 30^{\circ}$  사이의 영상 10개씩이고, 테스트 영상은 위로 향한 영상, 평행 시선의 영상, 아래 향한 영상 각 10개씩이다. 얼굴의 상하 각도는 과도하지 않고 대상이 자연스럽게 움직이는 정도로 설정하였다. 얼굴의 상하 변화는 좌우 변화와는 달리 각도 변화에 따라 영상의 변화가 크고 대칭성도 갖고 있지 않기 때문에, 큰 변화에 대한 보상은 어려울 것이라 판단했기 때문이다. 따라서 평균적으로 위로 약  $10^{\circ}$ , 아래로 약  $10^{\circ}$  정도로 향하는 영상들을 이용하였다. 실험에 사용된 영상의 예를 그림 14에 나타내었다.

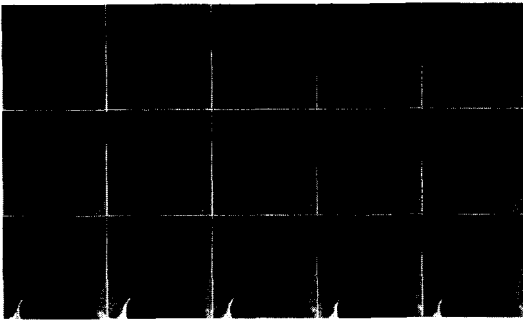


그림 14 실험 영상의 예(상, 중, 하)

표 3을 보면 이전의 실험에서 확인하였듯이 좌우 포즈 변환을 적용 시, 적용 전보다 세 가지 경우 모두 인식률이 크게 향상되었음을 볼 수 있다. 또한 좌우 포즈 변환을 적용하더라도 상하로 향한 영상은 평행한 시선의 영상에 비해 여전히 인식률이 떨어지는 것을 확인할 수 있다. 추가로 상하 포즈 변환을 적용했을 때는 조금 다른 결과를 볼 수 있다. 평행 시선의 경우 인식률이 약간 감소하기는 했지만 비슷한 결과를 볼 수 있었고, 아래로 향한 영상의 경우 오히려 인식률이 떨어졌으며, 위로 향한 영상의 경우에만 5.24%의 인식률 향상이 있었다. 아래로 향한 영상의 경우 그림 9에서 볼 수 있듯이, 정면 영상과 비교했을 때 코나 눈 등의 변화가 크기 때문에 변환을 통하여 오히려 왜곡이 더 커져 인식률이 떨어진 것으로 추측된다. 결국 실험 결과로부터, 시선이 위로 향한 영상의 경우에만 얼굴이 원통의 형태를 띠었다는 가정을 비교적 잘 따르는 것으로 볼 수 있으며, 따라서 이러한 경우에만 원통 모델을 이용하여 인식률의 향상이 가능한 것으로 보인다.

표 3 포즈 변환에 따른 인식률

방법 \ 시선	위로향함	평행 시선	아래로 향함
포즈변환 미적용시	59.05%	65.71%	48.10%
포즈변환 (좌우)	84.29%	95.71%	64.77%
포즈변환 (좌우+상하)	89.53%	95.24%	63.34%

표 4 결정 단계 융합에 의한 인식률 변화

방 법	적용 전	적용 후
확률값 이용 방법	83.33%	86.67%
발견적 방법(1)	83.33%	86.67%
발견적 방법(2)	83.33%	85.71%
발견적 방법(3)	83.33%	84.76%

결정 단계 융합으로 인한 인식률 변화를 표 4에 나타내었다. 적용 전의 인식률 83.33%는 좌우 포즈 변환 후 위로 향한 영상에 대해서만 상하 포즈 변환을 한 결과를 나타낸다. 결정 단계 융합 방법으로써 3장에서 기술한 실험적인 확률값을 이용한 방법과 그 외에 발견적(heuristic) 방법을 적용하여 실험을 하였다. 확률값은 실험적으로  $p_1=0.8333, p_2=0.0572, p_3=0.0444, p_4=0.0270, p_5=0.0064, \dots, p_{12}=0, \dots, p_{21}=0$ 을 획득하여 사용하였다 ( $p_1 + \dots + p_{21} = 1$ ). 융합 결과, 3.34%의 인식률 향상이 있었다. 또한, 발견적 방법으로서 다음과 같은 세 가지 방법을 적용해보았다. 첫째로, 앞에서 얻은 확률값을 점수로 사용하여 이전과는 달리 양 측 결과의 점수를 합산하는 방법이다. 둘째로, 실험을 통하여 획득한 점수 대신 인위적으로 순위별로 앞에서부터 6, 5, 4, 3, 2, 1점을 부여한 방법이다. 7번째 이상의 순위는 인식 결과에 영향을 끼치지 않는다는 가정 하에 0점을 부여하였다. 셋째로, 전과는 달리 확률값이나 점수를 고려하지 않고, 각각의 인식 결과인 벡터  $u, v$ 에서 인덱스(index)가 작은 부분( $u_1, v_1$ )부터 검색하여,  $u, v$  벡터 모두에 가장 먼저 나타나는 대상을 선택하는 것이다. 실험 결과를 보면 앞에서부터 각각 3.34%, 2.38%, 1.43%의 인식률 향상이 있었다. 결정 단계 융합 방법으로 어떤 방법을 택하느냐에 따라 인식률의 향상 정도가 다르기는 하지만, 전체적으로 인식률을 향상시킬 수 있음을 볼 수 있다.

#### 4. 결 론

본 연구에서는 원통 모델과 스테레오 카메라를 이용하여 입력 얼굴 영상의 포즈를 정규화 함으로써 대상의 포즈 변화에 강한 얼굴 인식 방법을 제안하였다. 입력으로써 단일 영상을 획득하는 경우 3차원 얼굴 모델보



다 비교적 간단하고 쉬운 원통 모델을 이용하여 좌우 포즈 변환을 실시함으로써, 대상의 좌우 포즈 변화에 강인한 인식 결과를 얻을 수 있다는 것을 보았다. 대상의 얼굴 변화 폭이  $-30^{\circ} \sim +30^{\circ}$  일 경우 좌우 포즈 변환을 통하여 인식률이 61.43%에서 94.76%로 향상되었다. 또한 3차원 얼굴 모델을 이용한 인식 결과와의 비교 실험을 통하여 3차원 얼굴 모델보다 간단하기는 하지만 양호한 인식 결과를 얻을 수 있다는 것을 보았다. 그러나 원통 모델을 이용하여 좌우 포즈 변환을 하더라도 동시에 상하 포즈 변화가 있는 대상에 대해서는 인식률이 떨어질 수밖에 없는 한계가 있다.

따라서 입력으로써 스테레오 영상을 획득할 경우, 얼굴의 상하 각도를 예측함으로써 좌우 포즈 변환에 추가하여 대상의 상하 포즈 변환 및 결정 단계 융합을 적용할 수 있었다. 물론 아래로 향한 영상의 경우 상하 포즈 변환의 효과를 얻을 수 없었지만, 위로 향한 영상의 경우 5.24%의 인식률 향상이 있음을 볼 수 있었다. 따라서 이를 이용하여 ATM 등의 고정된 장치에 활용할 수 있을 것이다. 또한 결정 단계 융합을 통하여 전체 인식률에서 최고 3.34%의 향상이 있음을 볼 수 있었다. 결국 본 논문에서 제안한 원통 모델과 스테레오 카메라를 이용한 얼굴 인식 방법을 적용함으로써, 비교적 간단한 방법으로 대상의 포즈 변화에 강인한 얼굴 인식 시스템을 구현할 수 있을 것이다.

참 고 문 헌

[1] D. J. Beymer, "Face recognition under varying pose," in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 556-761, Seattle, Washington, June 1994.

[2] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 84-91, Seattle, Washington, June 1994.

[3] F. J. Huang, Z. Zhou, H. Zhang, and T. Chen, "Pose invariant face recognition," in Proc. of IEEE Conf. on Automatic Face and Gesture Recognition, pp. 245-250, Grenoble, France, 2000.

[4] T. S. Jebara, "3D Pose estimation and normalization for face recognition," McGill University, 1996.

[5] S. Akamatsu, T. Sasaki, H. Fukumachi, and Y. Suenaga, "A robust face identification scheme - KL expansion of an invariant feature space," SPIE Proc., vol. 1607, pp. 71-84, Nov 1991.

[6] D. Graham and N. Allinson, "Face recognition from unfamiliar views: Subspace methods and pose dependency," in Proc. of IEEE Conf. on Automatic Face and Gesture Recognition, pp. 348-353, Nara, Japan, April 1998.

[7] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103-108, Jan 1990.

[8] M. Turk and A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.

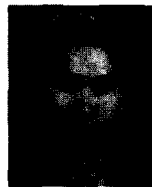
[9] M. H. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan 2002.

[10] T. Horprasert, Y. Yacoob, and L. Davis, "Computing 3-D head orientation from a monocular image sequence," in Proc. of IEEE Conf. on Automatic Face and Gesture Recognition, pp. 242-247, Killington, VT, Oct 1996.

[11] C. G. Feng, P. C. Yuen, and D. Q. Dai, "A novel method for face orientation determination in human face recognition," in Proc. of The Fourth Joint Conf. on Information Science, pp. 295-298, 1998.

[12] M. Xu and T. Akatsuka, "Detecting head pose from stereo image sequence for active face recognition," in Proc. of IEEE Conf. on Automatic Face and Gesture Recognition, pp. 82-87, Nara, Japan, April 1998.

[13] R. Hartley and A. Zisserman, Multiple View Geometry in computer vision, Cambridge University Press, 2000.



노진우  
1999년 2월 수원대학교 정보통신공학과 학사 졸업. 2003년 8월 고려대학교 전자공학과 석사 졸업. 2003년 8월~현재 삼성전자 연구원. 관심분야는 신호처리



홍정화  
1988년 2월 고려대학교 기계공학과 학사 졸업. 1990년 2월 고려대학교 기계공학과 석사 졸업. 1992년 6월 University of Wisconsin-Madison Engineering Mechanics 석사 졸업. 1996년 6월 Marquette Univ. Biomedical Eng. 박사 졸업. 1993년 7월~1996년 6월 Medical College of Wisconsin, Orthopaedic Research Division Research Associate. 1996년 6월~1999년 2월 삼성자동차 기술연구소 시험개발팀 책임연구원. 1999년 3월~2000년 4월 Genetal-Motors Technical Center, Vehicle Synthesis Analysis & Simulation Process Center, Automobile Safety Group Senior Researcher. 2000년 5월~2003년 2월 재할공

학연구센터 인체공학실책임연구원. 2003년 3월~현재 고려대학교 조교수. 관심분야는 제어계측공학, 메카트로닉스, 재활공학, 생체신호처리



고 한 석

1982년 5월 Canegie-Mellon Univ., Electrical Engineering 학사 졸업. 1986년 5월 Univ. of Maryland System Engineering 석사 졸업. 1988년 5월 Johns Hopkins Univ. Electrical Engineering 석사 졸업. 1992년 5월 Catholic Univ. of America, Electrical Engineering 박사 졸업. 1983년 9월~1995년 2월 White Oak 연구소, Principal Engineer. 1992년 9월~1995년 2월 Univ. of Maryland Part-time Asst. Professor. 1995년 3월~현재 고려대학교 전자컴퓨터공학과 교수. 관심분야는 전자 공학, 통신 공학, Data Fusion, Speech Recognition, Navigation and Tracking