

논문 2005-42SP-4-7

측면 포즈정규화를 통한 부분 영역을 이용한 포즈 변화에 강인한 얼굴 인식

(Face Recognition under Varying Pose using Local Area obtained by Side-view Pose Normalization)

안 병 두*, 고 한 석**

(Byeongdoo Ahn and Hanseok Ko)

요 약

본 논문에서는 측면 포즈 정규화를 통해 얻어진 부분영역을 이용해 대상의 포즈 변화에 강인한 얼굴인식 방법을 제안한다. 포즈변화에 강인한 얼굴인식을 위해 일반적으로 사용되는 방법인 포즈 정규화 방법은 포즈정규화과정 중에 가려져 보이지 않는 영역에 대한 정보를 가지고 있지 않기 때문에 문제가 발생하게 된다. 일반적으로는 보상을 통해 문제를 해결 하고 있지만, 보상에 의해 영상이 왜곡이 되거나 특징정보를 잃는 경우가 많다. 이런 문제를 해결하기 위해 깊이차가 큰 영역에서 주로 발생하는 왜곡을 줄이도록 정면이 아닌 측면으로의 정규화를 시도한다. 또한 정규화후 왜곡이 발생한 영역은 제거하고 왜곡이 발생하지 않은 영역만을 이용해 인식과정을 수행한다. 포즈가 좌우변화만 존재하는 경우와 상하변화도 존재하는 경우 두 가지 경우로 나누어 다루었으며 각각의 경우에 대해 실험을 통해 인식 성능의 향상을 확인하였다.

Abstract

This paper proposes a face recognition under varying poses using local area obtained by side-view pose normalization. General normalization methods for face recognition under varying pose have a problem with the information about invisible area of face. Generally this problem is solved by compensation, but there are many cases where the image is distorted or features lost due to compensation. To solve this problem, we normalize the face pose in side-view to reduce distortion that happens mainly in areas that have large depth variation. We only use undistorted area, removing the area that has been distorted by normalization. We consider two cases of yaw pose variation and pitch pose variation, and by experiments, we confirm the improvement of recognition performance.

Keywords : face recognition, face pose, local area, cylindrical model, pose invariant

I. 서 론

최근 IT 산업의 발전과 함께 보안이라는 영역에 대한 관심 또한 크게 증가하고 있다. 특히 유비쿼터스의 부각으로 인한 개인정보의 보호와 각국의 테러 위협으로 인한 출입국 보안 시스템까지 보안에 대한 필요성과 중요성에 대해 관심이 집중되고 있다. 그에 따라 보안을 위한 생체 인식에 대한 관심이 증대되고 있고 생체 인식 산업 역시 크게 각광받고 있다. 생체인식에는 크게 얼굴 인식, 지문 인식, 홍채 인식 등이 있으며 그 중

가장 각광 받는 분야는 얼굴인식이다. 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. 이러한 대상의 포즈변화에 대응하고자 얼굴 포즈를 정규화(normalization) 하는 방법^{[1][2][3]} 등이 제안되어왔다. 얼굴 포즈를 정규화 하는 방법은 입력 얼굴 영상을 정면 영상으로 바꾸어 인식하는 방법이다. 그러한 방법으로 어파인 변환(affine transform)이나^{[1][3]} 3차원 얼굴모델(3D face model)을 이용하는 방법이 있다^[2]. 어파인 변환은 대상의 깊이 정보를 고려하지 못하기 때문에,

* 학생회원, ** 정회원, 고려대학교 전자컴퓨터공학과
(Dept. of Electronics & Computer Engineering,
Korea University)

접수일자: 2004년12월22일, 수정완료일: 2005년1월31일

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 있기는 하지만, 3차원 얼굴 모델의 변형(deformation) 과정에서 계산량의 부하가 크며 전체적으로 그 과정이 복잡하다. 또한 대상의 자세로 인해 보이지 않는 영역에 대한 보상과정에서 왜곡이 발생하기도 한다. 따라서 깊이 정보를 고려한 3차원 변환의 효과를 가지면서 비교적 계산량도 적고, 변형과정중에 생기는 왜곡에 영향을 받지 않는 방법이 요구된다.

본 논문에서는 이러한 포즈 정규화 과정에서 발생하는 왜곡으로 인한 성능 저하를 제거하고 더 나은 성능을 얻기 위한 방법을 제안한다. 왜곡으로 인한 성능 저하를 막기 위해, 변형 과정에서 왜곡이 발생하는 보이지 않는 영역을 인식에 사용하지 않고, 보이는 부분 영역에 대해서만 정규화 과정을 거침으로써 더욱 정규화된 특징을 얻어 얼굴인식에 사용하도록 한다. 먼저 II장에서는 대상의 포즈가 좌우 변화만 있을 경우에 대해 어떻게 적용되는지를 기술하고, III장에서는 좌우 변화만이 아닌 상하 변화까지 포함되었을 경우에 대해 기술한다. 그리고 마지막으로 IV장에서 결론을 맺는다.

II. 좌우변환만을 고려할 경우

일반적으로 얼굴 인식에서 대상의 포즈를 다룰 때, 얼굴의 좌우 변화만을 고려하는 경우가 많다^{[4][5][6]}. 따라서 이 장에서는 포즈의 변화를 좌우 변화만이 있는 것으로 제한하여 다루고자 한다. 포즈의 예측과 정규화를 위하여 원통형 모델을 사용하였다.

1. 좌우 포즈 예측

좌우 포즈 변환을 위해 사용된 포즈 예측 모델은 그림 1에 나타내었듯이, 얼굴을 원통형이라 가정하여 얼굴의 좌우 폭과 두 눈의 위치를 이용한다. 대상의 포즈를 나타내는 두 눈 사이의 중간점이 얼굴 좌우 폭의 중심으로부터 떨어진 거리와 얼굴 폭의 반을 나타내는로부터 간단히 구해질 수 있으며 그 식은 아래와 같다.

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

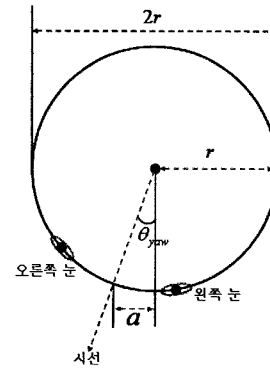


그림 1. 포즈 예측 모델

Fig. 1. Pose Estimation Model.

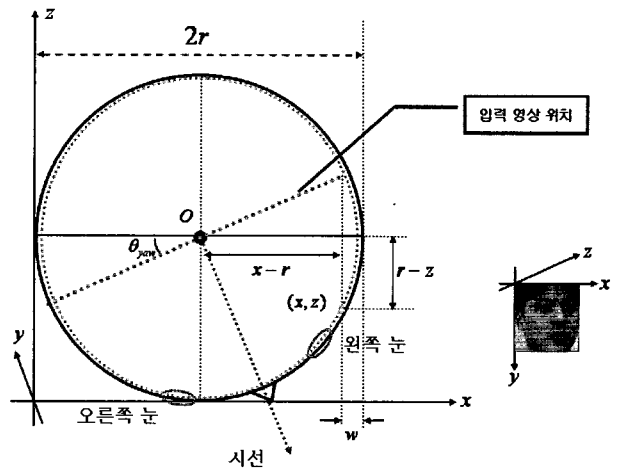


그림 2. 원통모형을 이용한 좌우 포즈 변환 구조도

Fig. 2. Structure of Yaw Pose Transform using Cylindrical Model.

2. 좌우 포즈 변환

대상이 바라보는 방향과 각도를 예측한 후, 얼굴 모양을 원통형이라 가정하여 포즈 정규화를 위한 변환을 실시하게 된다. 그 과정은 다음과 같으며, 구조는 그림 2에 나타내었다.

먼저 입력 얼굴 영상의 각 픽셀 위치 (x)에 해당하는 깊이 (z)를 구한다. 원통 모델로 가정하였으며 좌우 변환만을 고려함으로써 y 좌표에 따른 깊이 변화는 없으며 식 (2)로부터 식 (3)과 같은 깊이 정보를 얻을 수 있다.

$$(x-r)^2 + (r-z)^2 = r^2 \quad (2)$$

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

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



그림 3. 좌우 포즈 변환 결과
Fig. 3. Result of Yaw Pose Transform.

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

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 w만큼 제거하면 변환이 완료된다. 포즈 변환 결과는 그림 3에 나타내었다.

3. 부분영역검출

그림 3의 결과를 살펴보면 입력영상에서 보이던 영역에 대해서는 어느 정도 우리가 원하던 영상에 가까운 결과를 얻었음을 볼 수 있다. 그에 비해 보이지 않던 영역에 대해서는 예측을 통한 보상을 해주었음에도 우리가 원하던 영상과의 차이가 큼을 알 수 있다. 특히 코와 같이 깊이차가 큰 영역에 대해서는 그 왜곡이 더 커짐을 알 수 있다. 따라서 이러한 왜곡에 의해 인식률의 저하가 예상되며 이러한 문제를 해결하기 위해 두 가지 방법을 생각해 보게 된다. 첫 번째는 깊이차로 인한 왜곡을 줄이기 위하여 정면이 아닌 측면을 향하도록 정규화를 시켜주는 방법이다. 그러면 변환과정에 왜곡이 발생하였던 가려져 안보이던 영역은 축소되고 좋은 변환 결과를 보였던 드러나 보이던 영역을 주로 인식에 사용할 수 있게 될 것이다. 두 번째는 왜곡이 발생하는 영역은 제거하고 보이던 부분 영역에 대해서만 인식을 수행하는 것이다. 왜곡이 발생하는 영역을 제거함으로써 더 나은 성능을 얻을 수 있을 것이다.

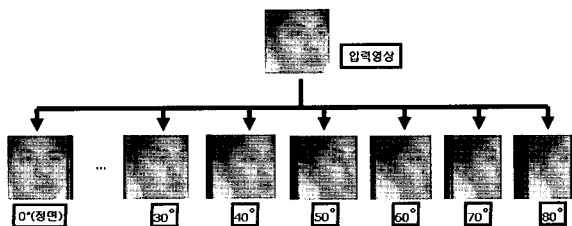


그림 4. 측면 정규화의 예
Fig. 4. Example of Side Normalization.

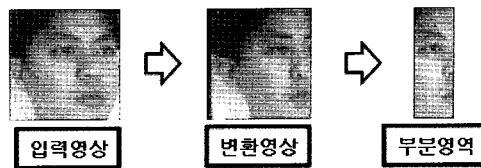


그림 5. 부분 영역 검출
Fig. 5. Detection of Local Area.

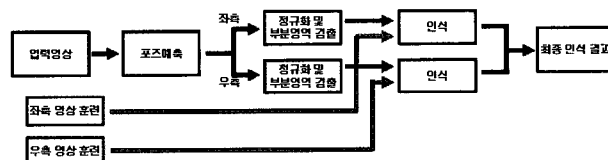


그림 6. 포즈 변화에 대한 제안된 얼굴인식 과정
Fig. 6. Process of Proposed Face Recognition System for Pose Variation.

그림 6에 전체 인식과정을 나타내었다. 카메라를 통해 들어온 영상은 원통 모델을 이용해 포즈가 예측 된다. 그리고 예측된 결과에 의해 정규화 되어지며 드러나 보이는 부분 영역에 대한 검출이 수행된다. 그 후 그 대상이 바라보고 있는 방향에 대해 훈련되어져 있는 결과를 이용해 인식과정을 수행한다.

4. 실험 및 결과

(1) 포즈 예측 방법의 실효성 판단

포즈 변환에 사용되는 포즈 예측 모델의 실효성을 판단하기 위해 0°~60° 까지 10° 간격으로 위치를 선정하고, 5명의 대상에게 그 지점을 바라보도록 요구하였으며, 각각 2번씩 획득하여 평균을 내었다. 획득되는 얼굴 영상의 폭은 80픽셀 전후이다. 촬영 각도와 예측 각도의 관계는 그림 5에 나타내었다. 예측 결과 0°~30°까지는 신뢰성이 있는 것으로 판단하였다. 30° 이후는 실제 움직이는 각도에 비해서 픽셀 값의 변화량이 적어서 오차가 크기 때문이다. 그리고 촬영 각도 10° 부근에서는 예측 값의 변화가 크게 나타났는데, 대상의 귀 모양 때문인 것으로 판단된다. 그림 6을 보면, 대상에 따라 얼굴에서 귀가 차지하는 폭이 다르기 때문에 동일한 각도에 대해서 예측 된 각도가 다를 수 있다.

(2) 포즈 변환된 부분영역을 이용한 인식률 향상평가

포즈 변환된 부분영역을 이용한 인식률 향상을 평가하기 위해 주 성분 분석법 (principal component analysis, PCA)을 사용하였다^{[7][8]}. 각 실험에 사용된 대상은 21명이며, 훈련 영상으로는 대상 당 10개, 테스트 영상으로는 대상 당 10개를 사용하였다.

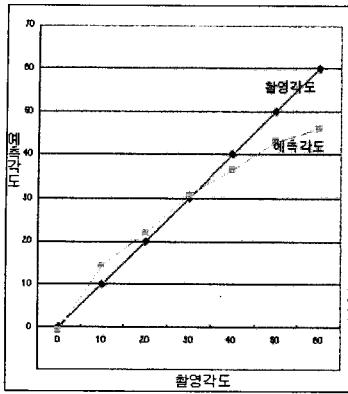


그림 7. 포즈 예측 결과
Fig. 7. Pose estimation results.

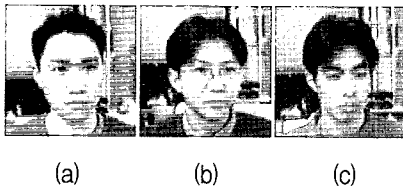


그림 8. 동일하게 10°를 바라 본 영상 (a)예측 각도 9.8°, (b)예측 각도 17.0°, (c)예측 각도 13.7°
Fig. 8. Images with same 10° view angle (a) estimated angle 9.8°, (b)estimated angle 17.0°, (c)estimate angle 13.7°.

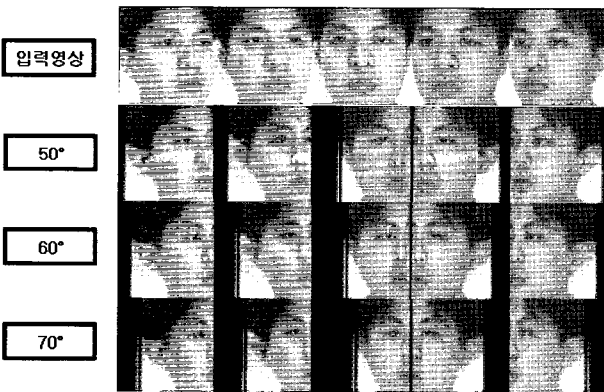


그림 9. 각도 별 정규화 변환의 예
Fig. 9. Example of Normalization Transform.

기존의 포즈 정규화 방법들은 대상을 정면을 바라본 방향으로 변환하여 인식을 수행하였다. 그러나 그림 3에서 보이듯이 코와 같이 깊이의 차가 존재하는 부분에서는 그 깊이차로 인해 정규화 과정에서 왜곡이 발생하게 된다. 따라서 정면이 아닌 약간의 측면에서 얻어진 영상이 오히려 깊이의 차로 인한 영상의 왜곡을 줄일 수 있을 것이라 예상하게 되어 그에 대한 실험을 해 보았다. 실험을 위해 측면으로 50°~80°까지 10° 간격으로 정규화를 시켜가며 인식률을 측정해 보았다.

정규화 각도 0°는 기존의 방법처럼 정면으로 정규화

표 1. 정규화 각도 별 인식률

Table 1. Recognition Rate for Normalization Degree.

정규화각도 \ 시선방향	좌측방향	우측방향
정규화 사용 안함	69.52%	67.62%
0° (정면으로 정규화)	96.18%	95.24%
50°	94.29%	93.33%
60°	97.14%	95.24%
70°	98.10%	98.10%
80°	98.10%	93.33%

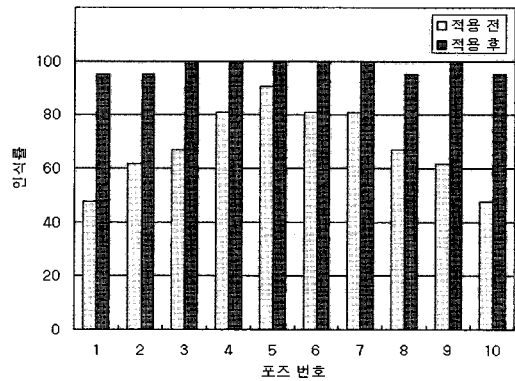


그림 10. 포즈 변화에 따른 인식률 비교
Fig. 10. Recognition rate with respect to pose.

한 경우이다. 표 1에서 보이듯 정규화 각도 70°에서 최고 인식률을 보였다. 정규화 각도가 70°일 때 가장 정규화 된 영상이 얻어지고, 따라서 특징 추출이 용이하기 때문으로 생각된다.

그림 10은 70도로 정규화 시켰을 경우의 각 번호별 인식률을 나타낸다. 번호 1번부터 10번까지는 각각의 대상이 좌측부터 우측으로 시선을 변화시켜가며 얻어진 영상이다. 실험 결과로부터 변환 전에는 정면영상에 비해 측면을 바라본 영상의 성능이 많이 떨어진데 비해, 변환 후에는 인식률이 어느 정도 일정하게 되었음을 알 수 있다. 따라서 제안된 알고리즘을 통해 시선이 정면에서 어느 정도 떨어진 영상의 경우에 대해서도 강인한 인식을 할 수 있음을 알 수 있다.

그림 6에서의 순서도에서, 제안된 알고리즘이 좌측을 향한 영상과 우측을 향한 영상을 각각 훈련 하고 인식에 이용한다는 것을 설명하였었다. 하지만 사람의 얼굴은 일반적으로 좌측과 우측이 대칭에 가까우므로 각각의 경우에 대해 따로 훈련하지 않고 한 번에 훈련하는 경우에 대해서도 생각해 볼 수 있게 된다. 예를 들어, 훈련영상에서 왼쪽을 향한 영상을 대칭변환을 시켜준 후에 오른쪽을 향한 영상과 한꺼번에 훈련을 시켜주고, 테스트 영상이 오른쪽을 향한 영상이면 훈련결과를 그대로 이용하고 왼쪽을 향한 영상이면 대칭변환을 시켜

표 2. 전체영상 훈련 시 정규화 각도 별 인식률
Table 2. Recognition Rate for Training Whole Images.

정규화각도 \ 시선방향	좌측방향	우측방향
50°	93.33%	95.24%
60°	95.24%	84.76%
70°	98.10%	84.76%
80°	94.29%	83.81%

준 후 훈련결과를 이용하는 것이다. 이 경우 각각의 경우에 대해서 훈련하고 인식을 수행하는 것보다 더 간단하면서도 적은 연산으로 인식과정이 수행될 것이다. 표 2는 그 결과를 나타내고 있다.

인식률이 각각에 대해 훈련한 경우보다 낮아진 것을 알 수 있었다. 연산량은 줄었지만 사람의 얼굴이 좌우가 완전한 대칭을 이루고 있는 것이 아니라 약간씩의 차이를 가지고 있기 때문에 인식률이 저하되었다고 생각된다.

III. 좌우 및 상하변환을 고려할 경우

앞 장에서는 얼굴의 포즈 변화에 대해서 좌우 변화만이 있는 것으로 제한하였다. 하지만 실제로 얼굴 영상을 획득할 때, 좌우 변화보다 상대적으로 작기는 하지만 어느 정도 상하 움직임이 존재하기 때문에 대상 얼굴의 상하 변화까지도 고려해야 한다. 좌우 포즈 변환을 적용하더라도 위아래로 향한 얼굴 영상들에 대한 보상이 없기 때문에, 위아래로 향한 영상들에 대한 인식률이 감소하는 것을 예상할 수 있었다. 따라서 상하 포즈 변화에 대한 보상을 통하여 인식률을 향상시킬 필요가 있다.

얼굴의 상하 포즈(pitch pose) 변화에 대한 보상을 하기 위해서는 우선 상하 포즈 예측이 필요하다. 이 논문에서는 상하포즈를 예측하기 위해 스테레오 기하학(stereo geometry)을 이용하고자 한다^{[9][10]}. 또한 스테레오 카메라 시스템을 이용하면 동시에 두 개의 영상을 얻을 수 있기 때문에, 결정 단계 융합(decision-level fusion)을 통하여 전체적으로 인식률을 향상시킬 수 있다는 장점이 있다.

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 변수(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 노멀 벡터(normal vector)를 이용하여 포즈를 예측한다.

내부 변수와 외부 변수로 이루어지는 투사 행렬(projection matrix) P, P' 를 획득하기 위해서, OpenCV 라이브러리의 캘리브레이션 필터(calibration filter)를 사용하였다.*

각 카메라의 투사 행렬 획득 후 양측 영상의 대응점으로부터 그 특징점의 3차원 좌표를 복원(reconstruction)하기 위하여 선형 삼각법(linear triangular method)를 이용하였다^[10]. 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, 호모지니어스(homogeneous) 좌표계로 표현하였다.

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

식 (5)로부터 아래의 3개의 식을 얻을 수 있다.

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

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

$$\begin{bmatrix} up_3^T - wp_1^T \\ vp_3^T - wp_2^T \\ u'p_3'^T - w'p_1'^T \\ v'p_3'^T - w'p_2'^T \end{bmatrix} M = TM = 0 \tag{7}$$

결국 특징점의 3차원 좌표 M 은 행렬 T 의 널벡터(null-vector)이며, 최소 제곱 근(least-squares solution)으로부터 구해질 수 있다.

대상의 포즈는, 이상의 방법으로 획득된 양 눈 끝과 입

* Intel Corp.에서 제공하는 OpenCV 라이브러리는 "http:// www.intel.com/research/mrl/research/opencv/"에서 구할 수 있다.

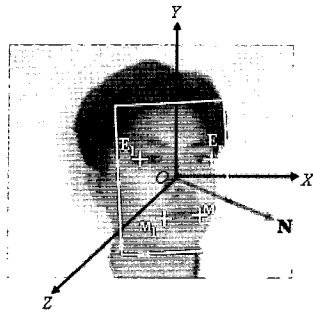


그림 11. 특징점의 3차원 좌표를 이용한 포즈 예측 모델
Fig. 11. Pose Estimation Model using 3D Coordinate of Feature Points.

끝의 4개의 특징점이 이루는 평면의 노말 벡터(normal vector)를 이용하여 구할 수 있으며, 그 개념을 그림 11에 나타내었다.

얼굴에 있는 4개의 특징점의 3차원 좌표로부터 포즈를 예측하는 과정은 다음과 같다. 우선 획득된 임의의 월드 좌표계에서의 3차원 좌표를 각각의 카메라 좌표계에서의 좌표로 변환한다.

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

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 법선 벡터를 구한다. 법선 벡터 N 은

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

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

결국 θ_{pitch} , θ_{yaw} 는 각각 법선 벡터와 X-O-Z 평면, Y-O-Z 평면이 이루는 각도와 같기 때문에 다음과 같이 구해진다.

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

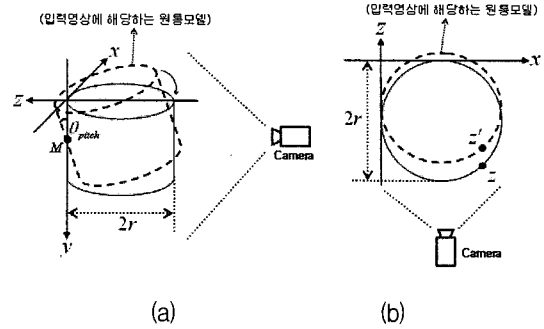


그림 12. 원통 모델을 이용한 상하 포즈 변환
(a) 측면에서 바라볼 때, (b) 위에서 바라볼 때
Fig. 12. Pitch Pose Transformation using Cylindrical Model. (a) View from Side, (b) View from Top.



그림 13. 상하 포즈 변환 결과
Fig. 13. Result of Pitch Pose Transform.

2. 상하 포즈 변환 과정

입력 영상의 상하 시선은 스테레오 시스템을 이용하여 θ_{pitch} 로 예측되었다. 따라서 입력 얼굴을 θ_{pitch} 만큼 회전된 원통형이라 가정하고, 그 모델에 맞춰 입력 영상의 각 픽셀 위치 (x, y) 에 해당하는 깊이 (z') 를 구한다. (그림 12)

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

r 은 원통의 반지름, z 는 정면인 원통 모델에서의 깊이, M 은 원통 회전의 기준점을 나타낸다. 이렇게 구한 3차원 좌표 (x, y, z') 를 θ_{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 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 매핑 시키고, 회전으로 인해 생성된 빈 픽셀 값은 양측 픽셀의 평균값으로 보충하고, 생성된 영상에서 정보가 없는 상하측 부분을 제거한다. 상하 포즈 변환 결과는 그림 11에 나타내었다. 원통 모델의 가정을 따라 대략적으로 정면에 가깝게 변환된 모습을 볼 수 있지만, 얼굴을 바라보는 각도에 따라 실제 눈, 코

의 변화가 크기 때문에 변환에 어려움이 있는 것으로 보인다.

3. Roll 변환 및 부분영역검출

좌우와 상하의 포즈 변화가 동시에 있을 경우는 단순히 좌우변환 후에 상하변환을 하는 순서의 단순 변환만으로는 정면에 가까운 영상을 얻을 수 없다. 대상이 바라보는 것은 측면의 상향 혹은 하향이기 때문에 비틀어져 있는 형태를 그대로 좌우 변환 후 상하변환을 시켜주게 되면 왜곡이 발생하게 된다. 따라서 그림 14 와 같이 롤(roll)의 개념을 적용하여 순차적으로 변환해 주는 것이 필요하다. 먼저 롤 변환을 통해 얼굴의 비스듬한 형태를 바로 세워준 후에 상하 그리고 좌우 변환을 거

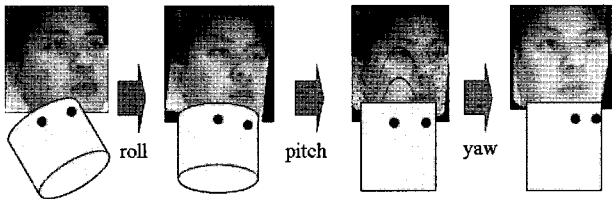


그림 14. 상하 포즈 변환에 대한 정규화 과정
Fig. 14. Normalization Process for Pitch Pose Variation.

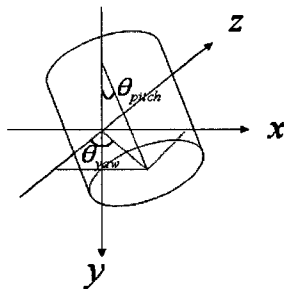


그림 15. 좌우 및 상하 변화에 대한 포즈 모델
Fig. 15. Pose Model for Yaw and Pitch Variation.

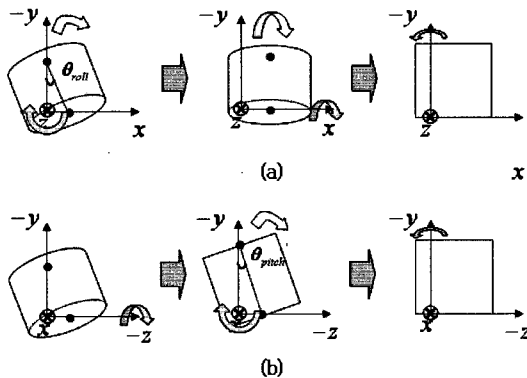


그림 16. 좌우 및 상하 변환 순서도
(a) 정면에서 바라볼 때 (b) 측면에서 바라볼 때
Fig. 16. Pose Transform for Yaw and Pitch Variation.
(a) View from Front, (b) View from Side.

쳐 얼굴의 형태를 정규화 시켜준다.

이 때 변환해 주어야 할 롤(roll), 피치(pitch), 여(yaw)의 각도는 그림 15의 구조에 의해 다음과 같은 식으로 구하게 된다.

$$\begin{aligned} \theta_{yaw}' &= \sin^{-1}(\cos\theta_{pitch} \cdot \sin\theta_{yaw}) \\ \theta_{pitch}' &= \sin^{-1}(\cos\theta_{yaw} \cdot \sin\theta_{pitch}) \\ \theta_{roll} &= \tan^{-1}\left(\frac{\sin\theta_{pitch} \cdot \sin\theta_{yaw}}{\cos\theta_{pitch}}\right) \end{aligned} \tag{14}$$

θ_{pitch} 와 θ_{yaw} 는 앞에서 설명한 상하 포즈 예측 모델에 의해 구해진 좌우 및 상하 포즈 각도이다.

이렇게 구해진 각도에 의해 롤 및 상하 포즈 변환을 하게 되면 영상은 θ_{yaw}' 만큼 평행한 방향으로 측면을 향한 얼굴영상이 된다. 그렇게 구해진 영상을 일정한 각도로 측면을 향하도록 정규화 시켜준 후 왜곡되지 않은 영역만을 추출하여 인식을 수행하게 된다.

4. 결정 단계 융합

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 17에 나타내었다. 양 측 카메라로부터의 각각의 인식 결과는 융합 과정을 통하여 최종적인 인식 결과를 도출하게 된다.

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.

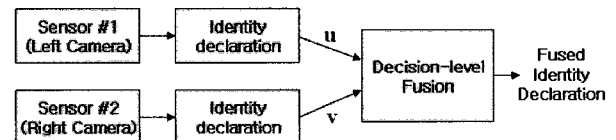


그림 17. 결정 단계 융합 구조도
Fig. 17. Structure of decision-level fusion.

$$\begin{aligned} \mathbf{u} &\in N^{M \times 1}, \mathbf{u} = [u_1 u_2 \cdots u_M]^T \\ \mathbf{v} &\in N^{M \times 1}, \mathbf{v} = [v_1 v_2 \cdots v_M]^T \end{aligned} \quad (15)$$

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

$$\begin{aligned} P(u_1 = H_i | H_i) &= p_1 \\ P(u_2 = H_i | H_i) &= p_2 \\ P(u_3 = H_i | H_i) &= p_3 \\ &\vdots \\ P(u_M = H_i | H_i) &= p_M \end{aligned} \quad (16)$$

$$\sum_{i=1}^M P(u_i = H_i | H_i) = p_1 + \cdots + p_M = 1 \quad (17)$$

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_l ($l=1, \dots, M$)은 다음과 같이 획득된다.

$$p_l = \frac{n_l}{T} \quad (18)$$

여기서 T 는 테스트 영상의 수를 의미하며, n_l 은 입력으로 들어온 대상이 결과 벡터 내에서 l 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 표현 될 수 있다.

$$P(H_i | \mathbf{u}, \mathbf{v}) = \frac{P(\mathbf{u}, \mathbf{v} | H_i) P(H_i)}{P(\mathbf{u}, \mathbf{v})} \quad (19)$$

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.

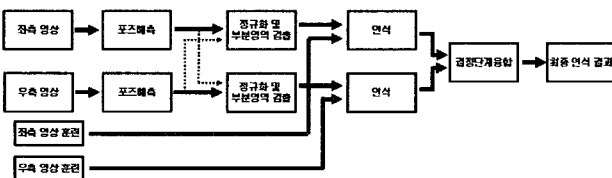


그림 18. 전체 얼굴인식 시스템 과정

Fig. 18. Process of Whole Face Recognition System.

$$\begin{aligned} P(H_i | \mathbf{u}, \mathbf{v}) &\propto P(\mathbf{u}, \mathbf{v} | H_i) \\ &= P(\mathbf{u} | H_i) \cdot P(\mathbf{v} | H_i) \\ &= P(u_j = H_i | H_i) \cdot P(v_k = H_i | H_i) \end{aligned} \quad (20)$$

\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 $\#i$ 에 해당하는 위치를 의미한다. 결국 인식 결과의 확률값을 이용하여 대상 $\#i$ 를 찾는 과정은 다음과 같이 표현된다.

$$\arg \max_i P(H_i | \mathbf{u}, \mathbf{v}) \quad (21)$$

지금까지 제안한 얼굴인식 과정은 그림 18 와 같이 요약할 수 있다.

5. 실험 및 결과

스테레오 시스템을 구축하기 위하여 CCD 기반의 웹 카메라(web camera) 2대를 사용하였다. 카메라의 해상도는 320×240 , 시야각은 62° 이며, 스테레오 카메라 세팅 구조도는 그림 19와 같다. 각 카메라가 바라보는 각도 θ 는 약 50cm 떨어진 대상의 행동 제한을 최소화 할 수 있는 각도로 설정하였다.

$$\theta = \tan^{-1}\left(\frac{L/2}{D}\right) = \tan^{-1}\left(\frac{10}{50}\right) \approx 11.3^\circ \quad (22)$$

각 카메라의 캘리브레이션을 위하여 OpenCV 라이브러리를 사용하였으며, 캘리브레이션 물체(calibration object)로는 그림 20과 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.

$$A = \begin{bmatrix} 362 & 0 & 171 \\ 0 & 361 & 95 \\ 0 & 0 & 1 \end{bmatrix} \quad A' = \begin{bmatrix} 364 & 0 & 184 \\ 0 & 362 & 135 \\ 0 & 0 & 1 \end{bmatrix} \quad (23)$$

임의의 월드 좌표계로부터 각 카메라 좌표계에 이르는 회전 행렬(rotation matrix) \mathbf{R}, \mathbf{R}' 와 이동 벡터(translation vector) \mathbf{t}, \mathbf{t}' 는 외부 변수로서, 두 카메라를 이용하여 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 10개씩이

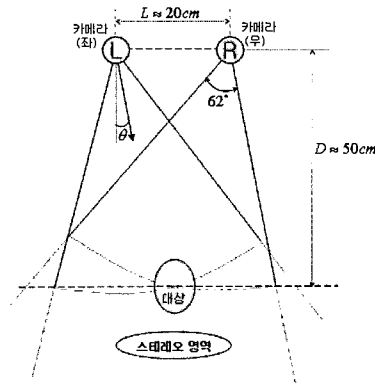


그림 19. 스테레오 카메라 세팅
Fig. 19. Stereo camera setting.

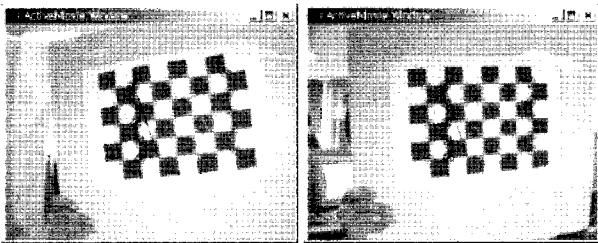


그림 20. 사용된 캘리브레이션 물체
Fig. 20. Calibration object used.

고, 테스트 영상은 위로 향한 영상, 평행 시선의 영상, 아래 향한 영상 각 10개씩이다. 얼굴의 상하 각도는 과도하지 않고 대상이 자연스럽게 움직이는 정도로 설정하였다. 따라서 평균적으로 위로 약 10°, 아래로 약 10° 정도로 향한 영상들을 이용하였다. 얼굴의 상하 변화는 좌우 변화와는 달리 각도 변화에 따라 영상의 변화가 크고, 대칭성도 갖고 있지 않기 때문에 그 이상의 각도는 고려하지 않았다.

표 3을 보면 이전의 실험에서 확인하였듯이 상하 포즈 변환을 적용 시, 적용 전보다 세 가지 경우 모두 인식률이 크게 향상되었음을 볼 수 있다. 각각의 시선방향에 대해 15~30%까지의 인식률 향상을 보였다. 하지만 위를 향한 시선과 아래를 향한 시선일 때에는 여전히 평행한 시선에 비해 낮은 인식률을 보임을 알 수 있다. 포즈 변환을 통해 어느 정도 보상이 되었으나 그 보상만으로는 부족함을 알 수 있다. 특히 아래를 향한 시선의 경우 인식률이 더 낮은 것으로 보았을 때, 상하 변환 시에 코영역에 존재하는 깊이 차에 의해 발생하는 왜곡에 대한 보상이 부족하기 때문으로 생각된다.

결정 단계 융합으로 인한 인식률 변화를 표 4에 나타내었다. 결정 단계 융합 방법으로써 4장에서 기술한 실험적인 확률값을 이용한 방법과 인덱스를 검색하는 방법을 적용하여 실험을 하였다. 확률값은 실험적으로 p_1

표 3. 포즈 변환에 대한 인식률

Table 3. Recognition Rate for Pose Transform.

	위를 향한	평행한 시선	아래를 향한
포즈변환 미 적용시	59.05%	65.71%	57.62%
포즈변환 적용시	78.10%	95.72%	71.43%

표 4. 결정 단계 융합 사용시 인식률

Table 4. Recognition Rate using Decision-level Fusion.

융합방법	시선			
	위를 향한	평행한 시선	아래를 향한	총
융합사용안함	78.10%	95.72%	71.43%	81.75%
확률값 사용	87.62%	99.05%	84.76%	90.48%
인덱스 사용	91.43%	99.05%	84.76%	91.75%

$=0.9810, p_2=0.0095, p_3=0.0095... , p_{21}=0$ 을 획득하여 사용하였다($p_1+\dots+p_{21}=1$). 확률값을 이용한 방법을 통해서 융합 결과, 8.7%의 인식률 향상이 있었다. 인덱스를 검색하는 방법은, 확률값은 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. 실험결과 확률값을 사용했을 경우보다 1.3%의 인식률 향상이 있었다. 특히 시선이 위를 향하고 있을 때 확률값을 이용한 방법보다 성능이 더 좋아 졌음을 알 수 있다. 결정 단계 융합 방법으로 어떤 방법을 택하느냐에 따라 인식률의 향상 정도가 다르기는 하지만, 전체적으로 인식률을 향상시킬 수 있음을 볼 수 있다.

IV. 결 론

본 논문에서는 부분영역을 이용한 포즈 변화에 강인한 얼굴인식 방법을 제안하였다. 기존 포즈 정규화 방법의 경우 보이지 않는 영역까지 정면으로 정규화 시키는 과정에서 왜곡이 발생하여 인식률이 저하되는 단점이 있었다. 이에 정면으로 정규화 시키지 않고 측면으로 정규화 시킴으로써 깊이의 차에 의한 왜곡을 최소화하며 작은 부분 영역에 많은 특징을 포함하도록 하였고 그렇게 얻어진 부분 영역만을 사용하여 인식을 수행함으로써 적은 연산량으로 인식속도를 개선하였다.

좌우 포즈 변화만 있을 경우 제안한 알고리즘에 의해 98%의 인식률이 측정 되어, 단순히 PCA만 사용했을 경우에 비해 약 30%의 인식률을 향상시킬 수 있었다. 측면으로 정규화 하였을 경우 70°로 정규화 하였을 때

가 인식률이 가장 높게 측정 되었으며 이는 70°에서 특징이 가장 정규화 되었기 때문으로 생각된다.

상하 포즈 변화까지 포함 되었을 경우 평행하게 바라보았을 경우는 95%로 비교적 높은 인식률을 보였으나 위나 아래를 바라보았을 경우는 각각 78%, 71%로 약간 낮은 인식률을 보이는 것을 알 수 있었다. 좌우의 포즈 변화에 대해 70°로 약간 측면을 향해 정규화 시키는 것이 가장 높은 인식률을 보이듯이 상하 포즈에 대해서도 오히려 약간의 상향, 혹은 하향에 대해서 높은 인식률을 보일 수 있다고 생각되며 이에 대해 추가적인 연구가 필요할 것이라고 생각된다.

또한 상하 변화에 대해서는 스테레오 카메라를 사용하였기 때문에 좌우 카메라 각각의 결과를 이용하여 결정단계융합(Decision Level Fusion)이 가능하다. 결정단계융합을 통하여 평행한 시선에 대해서는 99%, 위, 아래를 바라본 시선에 대해서는 각각 91%, 84%를 보여 결정단계 융합 전보다 훨씬 향상된 인식률을 보였다.

따라서, 본 논문이 제안하는 방법이 기존의 얼굴인식 시스템에 비해 효율적이며 포즈에 강인한 결과를 보여 주었고 본 논문에서 제안한 얼굴 인식 방법을 적용함으로써, 비교적 간단한 방법으로 대상의 포즈 변화에 강인한 얼굴 인식 시스템을 구현할 수 있을 것이다.

참고 문헌

- [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, Jun. 1994.
- [2] T. S. Jebara, "3D Pose Estimation and Normalization for Face Recognition," *McGill University*, May. 1996.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. on 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. 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.
- [10] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*, Cambridge University Press, 2000.

저자 소개



안 병 두(학생회원)
2003년 고려대학교 전기전자전파
공학부 학사 졸업.
2005년 고려대학교 전자컴퓨터
공학과 석사 졸업.
2005년~현재 삼성전자 연구원
<주관심분야: 신호처리>



고 한 석(정회원)
1982년 Canegie-Mellon Univ.,
전기 공학과 학사 졸업.
1986년 Maryland College Park,
시스템 공학과 석사 졸업.
1988년 존스 홉킨스 대학교 전기
공학과 석사 졸업.
1992년 Catholic Univ. of America, 전기공학과
박사 졸업
1983년~1995년 White Oak 연구소, 책임 연구원
1992년~1995년 Univ. of Maryland Baltimore
Contry, 조교수
1995년~현재 고려대학교 전자컴퓨터공학과 교수.
<주관심분야: 전자 공학, 통신 공학, Data Fusion,
Speech Recognition, Navigation and Tracking>