# 원통 모델과 스테레오 카메라를 이용한 포즈 변화에 강인한 얼굴인식

노 진 우, 안 병 두, *David Han, 고 한 석
고려대학교 전자공학과, *Johns Hopkins University
전화 : 02-922-8997 / 핸드폰 : 011-9939-4218

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

## 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 estimated object's pitch pose by stereo geometry. Also, since we have advantage that we can utilize two images acquired at the same time, we can increase overall recognition rate by decision-level fusion. By experiment, we confirmed that recognition rate could be increased using our methods.

## Ⅰ. 서론

얼굴 인식은 다른 인식 방법과 비교하였을 때 대상에게 특별한 동작을 요구하지 않는다는 장점을 갖고 있다. 하지만 그로 인해 대상의 포즈 변화에 민감하다. 이러한 대상의 포즈변화에 대응하고자 시각에 기반한 접근 방법과[1][2][3] 얼굴 포즈를 정규화(normalization) 하는 방법[1][4] 등이 제안되어왔다. 하지만 시각에 기반한 접근 방법은 요구되는 갤러리 영상의 수가 많고, 다수의 고유 공간을 구성해야 하기 때문에 훈련에 요구되는 샘플의 수도 많고 그만큼 훈련 부하도 커진다. 3차원 얼굴모델(3D face model)을 이용하여 포즈를 정규화 하는 방법은 실제 정면 얼굴에 가까운 영상을 얻을 수 있기는 하지만, 모델의 변형(deformation) 과정에서 계산량의 부하가 크며 전체적으로 그 과정이 복잡하다. 따라서 3차원 변환의 효과를 가지면서 비교적 계산량이 적은 방법이 요구된다.

본 논문에서는 3차원 모델로써 원통 모델을 사용하여, 비교적 간단한 방법으로 정면에 가까운 영상을 획득하여 인식에 이용한다. 입력으로 단일 영상을 획득하는 경우 포즈의 좌우 변환만을 통하여 포즈를 정규화 하는 방법을 제안하고, 스테레오 영상을 획득하는 경우 포즈의 상하 변환까지 적용하여 정규화 하는 방법 및 결정 단계 융합을 제안한다. 그리고 마지막으로 4장에서 결론을 맺는다.

## Ⅱ. 원통모델을 이용한 좌우 포즈 변환

일반적으로 얼굴 인식에서 대상의 포즈를 다룰 때, 얼굴의 좌우 변화만을 고려하는 경우가 많다[2][3][5]. 따라서 이 장에서는 포즈의 변화를 좌우 변화만이 있는 것으로 제한한다. 그 과정을 그림 1에 나타내었다.


그림 1. 단일 입력 영상의 경우 얼굴인식 과정

### 2.1 좌우 포즈 예측 방법

좌우 포즈 각도는 그림 2에 나타내었듯이 두 눈 사이의 중간점이 얼굴 좌우 폭의 중심으로부터 떨어진

거리 $a$와 얼굴 폭의 반을 나타내는 $r$을 이용하여 예측될 수 있으며, 그 식은 아래와 같다.

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

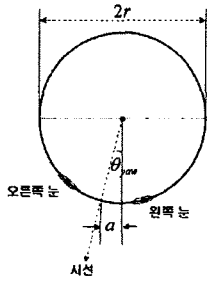


그림 2. 좌우 포즈 예측 모델

## 2.2 좌우 포즈 변환 과정

얼굴을 원통형이라 가정하여 포즈 변환 하는 과정은 다음과 같으며, 구조를 그림 3에 나타내었다. 먼저 입력 얼굴 영상의 각 픽셀 위치 $(x)$에 해당하는 깊이 $(z)$를 구한다.

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

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

$$x' = (x-r)cos\theta_{yaw} + (z-r)sin\theta_{yaw} - r \quad (3)$$

그 후 입력 얼굴 영상의 $(x)$좌표에 해당하는 픽셀값을 생성된 좌표 $(x')$에 매핑 시키고, 값이 없는 픽셀에는 양쪽 픽셀 값의 평균으로 보상하고, 양측을 $w$만큼 제거하면 변환이 완료된다. 포즈 변환 결과를 살펴보면(그림 4), 원통 모델의 가정을 따라 영상이 비선형적으로 늘거나 줄어들었음을 볼 수 있다.

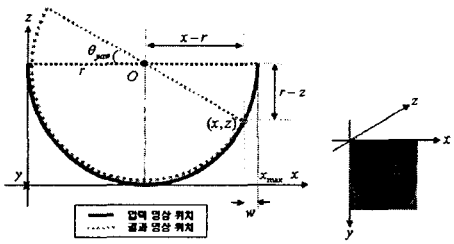


그림 3. 좌우 포즈 변환 구조도


그림 4. 좌우 포즈 변환 결과

## 2.3 실험 결과

(1) 좌우 포즈 변환으로 인한 인식률 향상 평가

얼굴 인식 방법으로 주성분 분석법(PCA)을 사용하였다[6]. 대상은 21명이며, 대상 당 훈련 영상 10개, 테스트 영상 10개를 사용하였다. 표 1을 보면, 좌우 포즈 변환으로 인해 인식률 향상이 있음을 확인할 수 있으며, 다루는 대상의 포즈 범위가 넓어질수록 변환으로 인한 효과가 더욱 증가함을 볼 수 있다.

표 1. 좌우 포즈 변환으로 인한 인식률 변화

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

(2) 3차원 얼굴 모델을 사용했을 때와의 비교 실험

3차원 얼굴 모델을 사용했을 때의 인식 결과와 비교하기 위하여, Jebara의 실험 결과와 비교하였다[4](표 2). 하지만 Jebara의 실험과는 달리 원통 모델 실험에서는 얼굴 및 눈의 검출을 하지 않고, 그것을 정확히 알고 있다고 가정하여 인식률을 얻었기 때문에 인식 결과를 직접적으로 비교할 수 없다. M. H. Yang et al.에 의한 조사에서 따르면 일반적으로 얼굴 검출은 약 90% 이상의 성능을 보이기 때문에[7] 이를 적용하여 비교한다. 따라서 최종적인 인식률은 75.60%로 볼 수 있으며, Jebara의 3차원 얼굴 모델보다 약 10%의 인식률 향상이 있게 된다. 비교적 간단한 모델인 원통 모델을 사용했을 때 인식률이 더욱 향상 된 이유는, Jebara가 계산량의 부하를 줄이기 위해 3차원 얼굴 모델의 변형 과정을 간략화 하여 수직 방향의 길이 조정만을 했기 때문인 것으로 보인다. 따라서 변형된 영상이 실제 정면 영상과 같은 모습을 보이기는 하지만, 인식률에 영향을 미치는 대상 간의 변화(variation)가 줄어들어 인식률이 감소한 것으로 보인다.

표 2. Jebara의 3차원 얼굴 모델과의 인식률 비교

| | 3차원 얼굴 모델 | 원통 모델 |
|---|---|---|
| 인식률 | 65.33% | 84.00% |

# Ⅲ. 스테레오 카메라를 이용한 상하 포즈 변환

앞 장에서는 좌우 변화만을 고려하였지만, 어느 정도 상하 움직임이 존재하기 때문에 상하 변화까지도 고려해야 한다. 얼굴의 상하 포즈를 예측하는 방법으로 단일 영상을 이용하는 방법들은 실제 얼굴 인식에 적용하기 어렵다. 따라서 스테레오 카메라를 이용하고자 한다[8][9]. 여기에 결정 단계 융합까지 적용한다.

## 3.1 상하 포즈 예측 방법

먼저 카메라 캘리브레이션을 통하여 내부 변수와 외부 변수를 획득한다. 이러한 정보와 양측 영상의 대응점을 이용하여 양쪽 눈 끝, 입 끝의 4개 특징점의 3차

원 좌표를 획득하고, 특징점들이 구성하는 평면의 노멀 벡터(normal vector) $N$을 이용하여 포즈를 예측한다. 포즈 예측 모델은 그림 5에 나타내었다.

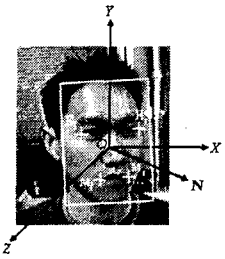


그림 5. 상하 포즈 예측 모델

카메라 캘리브레이션을 위하여 Intel OpenCV 라이브러리를 사용하였다. 스테레오 시스템을 이용하여 획득한 4개 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 하면, 평면의 방정식 $ax + by + cz + d = 0$ 으로부터 다음의 관계를 만들 수 있다.

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \mathbf{0} \tag{4}$$

식 (4)로부터 least squares solution을 이용하여 노멀 벡터를 구한다. 노멀 벡터 $N$은

$$N = a'\hat{x} + b'\hat{y} + c'\hat{z} \tag{5}$$

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

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

$$\theta_{pitch} = arc\sin\frac{b'}{\sqrt{b'^2 + c'^2}} \tag{6}$$

$$\theta_{yaw} = arc\sin\frac{a'}{\sqrt{a'^2 + c'^2}}$$

## 3.2 상하 포즈 변환 과정

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 6).

$$z' = (M - y)\sin\theta_{pitch} + (-r - \sqrt{x(2r - x)})\cos\theta_{pitch} \tag{7}$$

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

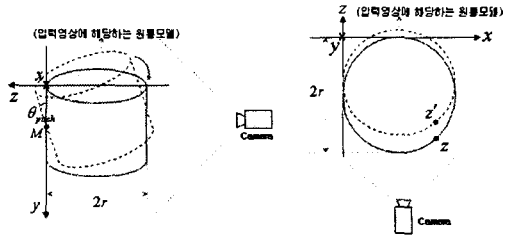


그림 6. 원통 모델을 이용한 상하 포즈 변환

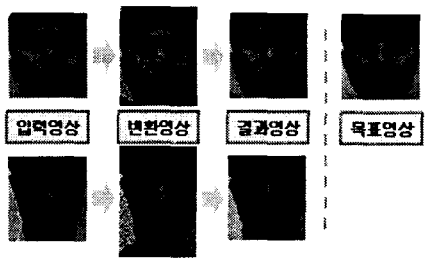


그림 7. 상하포즈 변환 결과

$$y' = -(M - y)\cos\theta_{pitch} + z'\sin\theta_{pitch} + M \tag{8}$$

그 후 입력 얼굴 영상의 $(x, y)$좌표에 해당하는 픽셀 값을 변환 된 $(x, y')$에 매핑 시키고, 회전으로 인해 생성된 빈 픽셀 값은 양측 픽셀의 평균값으로 보상하고, 생성된 영상에서 정보가 없는 상하측 부분을 제거한다. 상하 포즈 변환 결과는 그림 7에 나타내었다.

## 3.3 결정 단계 융합

스테레오 카메라를 사용하면 동일한 대상에 대해서 다른 포즈의 영상을 동시에 획득할 수 있기 때문에, 결정 단계 융합을 통하여 인식률 향상을 꾀할 수 있다.

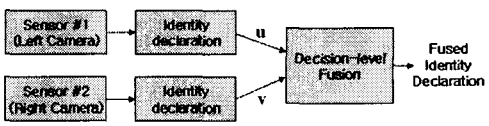


그림 8. 결정 단계 융합 구조도

융합 과정의 입력 $u$, $v$는 각각 M×1의 벡터로서(M은 대상의 수), 입력 영상에서 추출된 PCA 특징과 데이터베이스 내의 각 대상의 PCA 특징과의 유클리드 거리가 작은 순서대로, 그 대상을 나타내는 번호를 $u_1$부터 $u_M$에(또는 $v_1$부터 $v_M$에) 할당한다.

$$u \in N^{M\times 1},\ u = [u_1\ u_2\ \cdots\ u_M]^T \tag{9}$$

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

$$\sum_{i=1}^{M} P(u_i = H_i | H_i) = p_1 + \cdots + p_M = 1 \tag{10}$$

$H_i$는 관측된 존재(entity)가 대상 #$i$인 것을 나타낸

다. 융합을 통하여 각 대상에 대한 결합 확률을 구하고, 그 값이 가장 큰 대상이 최종 인식 대상으로 한다. 물론 두 명의 대상이 같은 값을 갖는 경우, 보다 정면에 가까운 영상을 입력으로 갖는 쪽의 결과를 따른다.

$$\arg\max_{i} P(H_i \mid \mathbf{u}, \mathbf{v}) \tag{11}$$

### 3.4 실험 결과

(1) 상하 포즈 변환으로 인한 인식률 향상 평가

상하 포즈 변환을 적용한 인식률을 표 3에 나타내었다. 대상은 21명이며, 훈련 영상은 대상 당 평행 시선의 -30°~+30° 사이의 영상 10개, 테스트 영상은 상향, 평행, 하향 영상 각 10개씩이다. 상하 각도는 과도하지 않게 설정하였다. 참고로 이번에는 훈련 영상과 테스트 영상이 다른 날에 촬영된(duplicate) 경우를 고려하여 표 1과 비교하였을 때 인식률이 전체적으로 낮다. 이와 관련한 내용은 P. J. Phillips et al.에서 참고할 수 있다[10]. 결과를 보면 좌우 포즈 변환을 통하여 인식률이 크게 향상되었지만, 상하로 향한 영상은 평행한 시선의 영상에 비해 인식률이 떨어진다. 상하 포즈 변환을 적용했을 때는 위로 향한 영상의 경우에만 10%의 인식률 향상이 있었으며, 아래로 향한 영상의 경우 오히려 인식률이 떨어졌다. 그림 7에서 볼 수 있듯이, 정면 영상과 비교했을 때 코나 눈 등의 변화가 크기 때문에 변환을 통하여 오히려 왜곡이 더 커져 인식률이 떨어진 것으로 추측된다. 결국 시선이 위로 향한 영상의 경우에만 얼굴이 원통의 형태를 띠었다는 가정을 따라 인식률의 향상이 가능한 것으로 보인다.

표 3. 포즈 변환에 따른 인식률

| 방법 \ 시선 | 위로향함 | 평행 시선 | 아래로 향함 |
|---|---|---|---|
| 포즈변환 미적용시 | 30.48% | 48.10% | 30.48% |
| 포즈변환 (좌우) | 45.72% | 70.96% | 59.05% |
| 포즈변환 (좌우+상하) | 55.72% | 69.53% | 44.76% |

(2) 결정단계융합으로 인한 인식률 향상 평가

결정 단계 융합으로 인한 인식률 변화를 표 4에 나타내었다. 적용 전의 인식률 61.90%는 좌우 포즈 변환 후 위로 향한 영상에 대해서만 상하 포즈 변환을 한 결과를 나타낸다. 확률값은 실험적으로 $p_1$=0.6190, $p_2$=0.0937, $p_3$=0.0635, ... , $p_{21}$=0 을 획득하여 사용하였다 ($p_1+\cdots+p_{21}=1$). 융합 결과 3.5%의 인식률 향상이 있었다.

표 4. 결정 단계 융합에 의한 인식률 변화

| | 적용 전 | 적용 후 |
|---|---|---|
| 인식률 | 61.90% | 65.40% |

## Ⅳ. 결론

본 연구에서는 원통 모델과 스테레오 카메라를 이용하여 입력 얼굴 영상의 포즈를 정규화 함으로써 대상의 포즈 변화에 강인한 얼굴 인식 방법을 제안하였다.

입력으로써 단일 영상을 획득하는 경우 3차원 얼굴 모델보다 비교적 간단하고 쉬운 원통 모델을 이용하여 좌우 포즈 변환을 실시함으로써, 대상의 좌우 포즈 변화에 강인한 인식 결과를 얻을 수 있다는 것을 보았다. 그러나 상하 포즈 변화에 대해서는 인식률이 떨어질 수밖에 없는 한계가 있다. 따라서 입력으로써 스테레오 영상을 획득할 경우, 좌우 포즈 변환에 추가하여 대상의 상하 포즈 변환 및 결정 단계 융합을 적용할 수 있었다. 상하 포즈 변환을 적용했을 때 위로 향한 영상의 경우 10%의 인식률 향상이 있었으며, 결정 단계 융합을 통하여 전체 인식률에서 3.5%의 향상이 있었다. 결국 본 논문에서 제안한 포즈 정규화 방법을 이용하여, 비교적 간단한 방법으로 대상의 포즈 변화에 강인한 얼굴인식 시스템을 구현할 수 있을 것이다.

## 참고문헌


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

[2] A. Pentland, B. Moghanddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *in Proc. of IEEE Conf. on CVPR*, 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 AFGR*, pp. 245-250, Grenoble, France, 2000.

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

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

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

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

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

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

[10] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. PAMI*, vol. 22, no. 10, pp. 1090-1104, Oct 2000.