

2 차원 영상의 얼굴 각도와 위치 판정

김우원, 황진규, *이상윤
연세대학교 생체인식협동과정, *연세대학교 전기전자공학과
woowon@yonsei.ac.kr, winispirit@yonsei.ac.kr, *syleee@yonsei.ac.kr

Estimation of Head Angle and Position from 2D images

Woowon Kim, Jinkyu Hwang, *Sangyoung Lee
Yonsei University Biometrics, *Yonsei University Electrical Engineering

요 약

얼굴의 각도와 위치는 많은 정보를 포함하고 있어 다양한 분야에서 응용이 가능하다. 본 논문에서는 하나의 2 차원 영상만으로 사람 얼굴의 3 차원 각도와 위치를 파악하는 시스템을 제안한다. 얼굴 특징 점 검출 기술, 포즈 판정 기술, 오류 안정화 기술 등을 단계별로 설명하며, 비디오 데이터베이스를 통한 실험으로 시스템의 성능을 분석하였다. 오류는 5 도 안팎의 적은 수치이며 개인용 컴퓨터에서 실시간 수행이 가능하다.

1. 서론

사람의 얼굴 각도와 위치는 많은 정보를 내포하고 있다. 얼굴로 사물을 쳐다 보는 것으로 의도를 나타낼 수도 있고, 고개의 끄덕임 등을 통해 의사 표현을 할 수도 있다. 얼굴 각도와 위치 정보는 인간과 컴퓨터의 상호작용에서도 응용된다. 현재 응용된 예로는 얼굴의 각도를 사용한 컴퓨터 마우스 커서의 제어 [1], 자동차 운전자의 의도와 졸음 정도 판단 [5], 사용자 머리의 3 차원 위치를 추적할 통한 2 차원 화면에서의 입체감있는 디스플레이 등이 있다. 이 외에도 게임이나 TV 리모콘에서 다양한 제어, 다각도의 사용자 바이오 인식 또한 미래에 가능할 수 있다 [6].

얼굴의 포즈(3 차원 각도와 3 차원 위치)의 알고리즘은 크게 두 분야로 나눌 수 있는데, 3 차원 스캐너나 다중 카메라를 이용한 3 차원 정보를 통한 판정방법과 하나의 2 차원 영상만으로 3 차원 정보를 판정하는 방법이 있다. 3 차원 정보를 이용할 경우 정확성이 보장되지만, 기계의 가격과 복잡한 설정 때문에 응용이 제한적이다. 하나의 2 차원 영상을 이용하는 알고리즘에서는 현재 영상 추적을 기반으로 하는데, 이는 초기 설정시의 정확성이 요구되며, 실행 시간이 지나면서 추적되는 객체와 추적하려는 객체 간 오차가 생기며 오류가 점차 증가한다는 문제점이 있다.

본 논문에서는 하나의 2 차원 영상만으로 포즈 판정을 하며 영상 추적에 기반하지 않는 방법을 제시한다. 본 논문에서는 2 차원 영상의 얼굴에서 특징 점을 찾아, 특징 점의 2 차원 위치와 학습된 얼굴의 3 차원 정보와 비교하여 객체의 3 차원 포즈를 판정한다. 또한, 판정이 잘 못 이루어졌을 때 학습된 얼굴을 사용자에게 피팅하여 오류를 안정화 방법에 대해서도 소개한다.

본 논문의 구성은 다음과 같다. 2 절에서는 2 차원 영상에서

얼굴 특징 점을 찾는 기술, 3 절에서는 특징 점들로부터 포즈를 찾는 기술, 4 절에서는 오류를 안정화하는 기술을 설명한다. 5 절에서는 본 논문에서 소개한 기술들을 취합하여 비디오 데이터베이스에서의 성능을 실험한다. 6 절에서는 본 논문을 마무리한다.

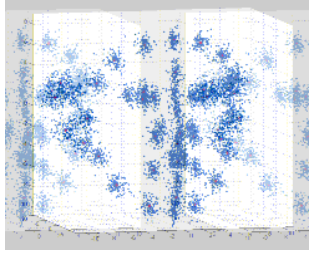
2. 얼굴 특징 점 검출

특징 점 검출의 목적은 은 영상 상에서 얼굴의 특정 부분(눈동자, 입 끝 등)을 찾아 영상의 좌표로 반환하는 것이다. 본 논문에서는 특징 점 검출에 조명과 표정변화에도 강인하면서도 빠른 속도로 다수의 점을 검출할 수 있는 MCT-based AdaBoost (Modified Census Transform-based AdaBoost)를 사용하였다. MCT-based AdaBoost 는 특징 검출시 픽셀 주변에 따라 특징을 정하게 됨으로 일반적으로 사용되는 Haar-like feature AdaBoost 의 취약점인 오수락률(false-acceptance rate)을 개선하여 다른 AdaBoost 알고리즘 보다 높은 정확도를 보인다 [3]. 또한, 속도 개선으로 모바일 환경에서 10fps 의 속도를 나타낸다. 원 논문에서는 MCT-based AdaBoost 를 전면 얼굴 영역과 눈동자의 위치를 찾는 데에 사용하였지만, 우리는 같은 방법으로 코의 영역과 입의 양 끝을 찾을 수 있도록 학습하였다. 본 논문에서 검출하는 얼굴 특징 점은 양 눈의 점, 코, 양 입 끝 등의 5 점이다. 검출이 되기 쉬운 점을 선택함으로써 포즈 판정 단계에서 오류를 줄일 수 있다. MCT-based AdaBoost 를 사용한 5 점의 얼굴 특징 점 검출은 정지영상 데이터베이스를 통한 실험에서 다양한 포즈, 개인, 조명, 표정 변화에 강인한 성능을 나타내었다.

3.3 차원 포즈 판정

2 절의 단계에서 찾은 특징 점들로부터 포즈를 계산하기 앞서, 점들의 3 차원 위치가 정의될 필요가 있다. 우리는 *Cyberware™* 스캐너를 이용하여 얻은 100 명의 한국 남녀의 3D 스캔 데이터에서 선택된 점들의 3 차원 위치를 찾아 그 분포를 모델화하였다. 그림 1 은 모델화된 분포를 나타낸다. 우리는 이 모델 분포의 평균으로 이루어진 얼굴 모델을 만들어 3 차원 포즈 판정에 사용하였다.

그림 1. 스캔 데이터의 모델 분포
Fig. 1. Scan data model point distribution



포즈를 판정하기 위해 3 차원 좌표들과 입력 영상에서 검출된 2 차원 좌표들을 x 와 y 축의 비례공식으로 나타낼 수 있는데, 이를 행렬형으로 다음과 같이 나타낼 수 있다.

$$\begin{matrix} AT - B \\ T - AB \end{matrix} \quad A = \begin{bmatrix} M_{1x} & M_{1y} & M_{1z} & 1 \\ M_{2x} & M_{2y} & M_{2z} & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad T = \begin{bmatrix} i_x & j_x & 0 \\ i_y & j_y & 0 \\ i_z & j_z & 0 \\ -X & Y & 1 \end{bmatrix} \quad B = \begin{bmatrix} m_x & m_y & 1 \\ m_{2x} & m_{2y} & 1 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

위 식에서 A 행렬은 3 차원 모델의 좌표를 나타내며, 구성요소인 M_{1x} 은 각 특징점의 3 차원 축 좌표를 나타낸다. B 행렬은 2 차원 영상의 좌표를 나타내며, 구성요소인 m_{1x} 는 마찬가지로 각 특징점의 2 차원 좌표를 나타낸다. $AT=B$ 식에서 변환행렬 T 는 두 좌표 시스템간의 관계, 즉 얼굴의 각도와 위치를 나타낸다. A 행렬의 의사역행렬(pseudo-inverse)을 취하여 우측변의 B 행렬과 곱하면 변환행렬 T 를 구할 수 있다. 사용되는 3 차원 모델은 일정하게 지켜지기 때문에, 의사역행렬은 초기에 한 번만 계산되면 이후 입력 영상이나 다른 사용에도 계속 같은 행렬을 이용할 수 있다. (단, 4 절의 오류안정화 과정에서 모델이 변할 때 의사역행렬이 다시 구해질 필요가 있다.)

T 행렬의 구성요소에서 i_x, i_y, i_z 는 두 좌표의 x 축의 비례관계를 나타내는 벡터 \mathbf{i} , j_x, j_y, j_z 는 y 축의 비례관계의 벡터 \mathbf{j} 를 나타낸다. 비례관계를 나타내는 두 벡터는 각도 판정에 필요한 정보를 내포하고 있다. 이 정보를 얻기 위해 두 벡터로부터 직교행렬(orthonormality)의 성질을 갖고 있는 회전형렬로 전환할 수 있다. 벡터 \mathbf{i} 와 \mathbf{j} 를 단위행렬로 만들고, 두 벡터의 외적을 z 축의 비례관계 \mathbf{k} 로 두면 직교행렬의 성질이 성립된다. 생성된 행렬은 곧 3 차원 좌표와 2 차원 좌표의 회전형렬, 즉 영상에서 보이는 얼굴의 각도가 얻어진다.

$$\mathbf{k} = \mathbf{i} \times \mathbf{j} \quad Z = \frac{f}{(\text{norm}(\mathbf{i}) + \text{norm}(\mathbf{j})) / 2}$$

$$\text{Rotation} = \begin{bmatrix} i_x & i_y & i_z \\ j_x & j_y & j_z \\ k_x & k_y & k_z \end{bmatrix} \quad \text{Translation} = \begin{bmatrix} -X \\ Y \\ Z \end{bmatrix}$$

T 행렬의 다른 구성요소인 X 와 Y 는 전 단계에서 \mathbf{i} 와 \mathbf{j} 를 단위행렬로 만들기 위해 취한 각 벡터의 크기(norm)로 나누면, 3 차원 좌표의 원점과 2 차원 좌표와의 거리, 즉 영상에서 보이는 얼굴의 위치를 나타낸다. 두 norm 의 평균을 영상의 초점 거리 (focal length)와 나누면 Z , 즉 영상에서 보이는 얼굴 위치의 깊이의 값을 구할 수 있다.

4. 오류 안정화

3 절에서 소개한 포즈 판정은 의사역행렬을 사용하므로 발생할 수 있는 오차를 최소화하는 성질을 가지고 있다. 따라서 사용자의 얼굴이 생성된 모델과 차이가 있어도 대부분의 경우 편차 10 도 가량의 오류 내에 각도가 판정되었다. 하지만 사용자의 얼굴이 모델과 많은 차이가 날 경우 오차가 편차 15 도 가량으로 늘어나는 경우가 발생했다. 이런 경우의 오차를 줄이기 위해 사용자에게 맞추어 스캐너 데이터를 조절해주는 비디오 영상의 경우 다음 프레임의 오류를 안정화할 수 있다.

데이터의 조절은 x 와 y 축, 정면 얼굴에서 봤을 때의 왼쪽-오른쪽과 위-아래, 에서 진행된다. 데이터를 조절해주기 위해서는 먼저 정면을 바라보고 있는 영상이 필요하다. 이런 영상은 사용자에게 정면을 바라보라고 요구 할 수도 있지만, 자동으로 스캐너 데이터를 이용하면서 각도 판정 중 미리 설정된 한계점 이내 정면 영상이 검출되었을 경우 적용 되게 하였다. 정면 영상이 검출되었을 때 눈의 위치를 기준으로 3 차원 모델의 입과 코의 위치를 변환하여 사용자에게 취합된 모델로 보다 편차 5 도 가량의 더 정확한 결과가 나타났다.

5. 성능 실험

객관적인 실험을 위해 우리는 공개 데이터베이스인 BU dataset 을 사용하였다 [4]. BU dataset 은 5 명의 개인이 각 9 개의 자유롭게 변환하는 포즈를 각 5 초 가량의 비디오들로 모아놓은 데이터베이스이다. 각 비디오는 320x240 해상도로 초당 30 프레임으로 200 프레임으로 구성되어 있다. 또한 이 데이터베이스는 실험자들의 머리에 각도측정기가 부착되어 실제 값과의 성능 비교가 가능하다. 그림 2 는 제안된 포즈 판정 기술을 BU dataset 의 세 비디오에 적용한 그래프이다.

그림 2. 오류 안정화 이전의 3 비디오 에서의 각도 판정 결과
Fig. 2. Angle output on three videos from BU dataset before model update

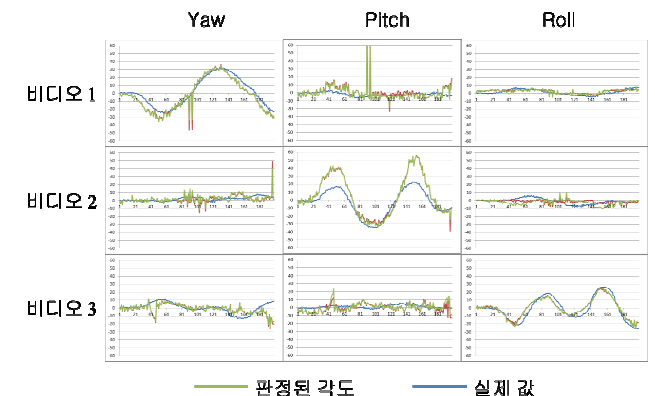


그림 2 에서 yaw 는 얼굴의 좌우변환, pitch 는 얼굴의 상하 변환, roll 은 얼굴의 기울임을 나타낸다. 그림 2 에서 사용된 비디오 1 과 비디오 3 에서는 얼굴의 x 축과 z 축의 변환에 강인한 성능을 보였다. 비디오 2 에서의 y 축 변환은 사용자의 코의 높이 차이로 인해 다소 차이를 보였다. 이를 해결하기 위하여 비디오 2 의 정면 얼굴에 오류 안정화를 적용하였고, 결과는 그림 3 과 같다. 안정화된 모델의 사용으로 오류의 범위가 줄어든 것을 확인할 수 있다.

그림 3. 안정화 적용 후의 각도 판정 결과
Fig. 3. Angle output after model update

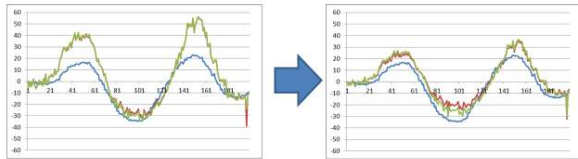


표 1 은 오류 안정화 알고리즘으로 인한 오류 개선을 나타낸다. 제안된 알고리즘을 BU dataset 의 전 비디오에 적용된 결과를 각 축 별 평균 오차 범위로 나타내었다.

표 1. BU dataset 에 적용된 알고리즘의 오류 도표

Table. 1. Error table of the proposed method on BU dataset

	안정화 이전	안정화 이후
Yaw	5.37	3.37
Pitch	9.21	4.80
Roll	3.58	3.44

(단위: 도)

모든 실험은 BU dataset 의 다양한 사용자와 조명의 변환에서 이루어졌으며, 데이터베이스의 5 실험자당 9 시퀀스의 200 프레임인 총 9000 개의 프레임의 평균 오류를 나타내었다. 실험에서의 평균 오류는 5 도 이내로 나타났으며, Intel i5 2G RAM 사양의 컴퓨터에서 50 fps 로 진행되었다.

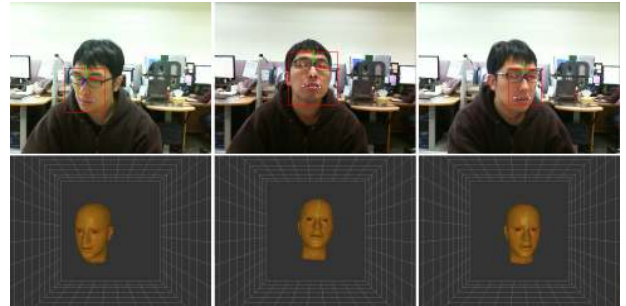
5. 결론

본 논문에서는 하나의 영상 카메라만으로 사용자의 얼굴을 찾아 그 얼굴의 위치와 각도를 알아내는 시스템을 소개하였다. 제안된 방법은 배경, 명암 등 환경의 변화에 강인하였으며, 개인에 따른 차이에도 오류의 안정화를 통해 더욱 좋은 성능을 나타낼 수 있었다. 비디오 데이터베이스의 성능 테스트에서는 실시간 적용의 가능성을 확인하였다.

향후 우리는 제안된 방법을 더욱 개선해 더 다양한 분야의 어플리케이션에 적용 가능한 시스템을 개발할 것이다. 오류 안정화 알고리즘의 개선과 추적 알고리즘과의 취합을 통하여 속도를 개선 하며 오류를 더욱 안정화 할 수 있을 것이다.

그림 4. 얼굴 포즈판정 정보를 이용하는 3D 데모

Fig. 4. Demo utilizing head pose estimation information



감사의 글

이 논문은 2011 년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2011-0016302).

참고 문헌

- [1] Betke, M., J. Gips, et al. (2002). "The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities." Neural Systems and Rehabilitation Engineering, IEEE Transactions on 10(1): 1-10.
- [2] Dementhon, D. F. and L. S. Davis (1995). "Model-based object pose in 25 lines of code." International Journal of Computer Vision 15(1): 123-141.
- [3] Jun, B. and D. Kim (2007). "Robust real-time face detection using face certainty map." Advances in Biometrics: 29-38.
- [4] La Cascia, M., S. Sclaroff, et al. (2000). "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(4): 322-336.
- [5] Murphy-Chutorian, E., A. Doshi, et al. (2007). Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation, IEEE.
- [6] Murphy-Chutorian, E. and M. M. Trivedi (2009). "Head pose estimation in computer vision: A survey." Pattern Analysis and Machine Intelligence, IEEE Transactions on 31(4): 607-626.