

스테레오 영상을 이용한 얼굴 포즈 추정

소인미*, 강선경*, 김영운*, 이지근*, 정성태**

Face Pose Estimation using Stereo Image

In-Mi So*, Sun-Kyung Kang*, Young-Un Kim*, Chi-Geun Lee*, Sung-Tae Jung**

요약

본 논문에서는 두 대의 카메라 영상으로부터 얼굴의 포즈를 추정하는 방법을 제안한다. 제안된 방법은 먼저 두 얼굴 영상으로부터 대응되는 눈썹, 눈, 입의 특징점을 추출한 다음, 스테레오 비전의 삼각법에 의해 특징점에 대한 3차원 위치를 계산한다. 그 다음에는 특징점으로 부터 삼각형을 생성하고 그 삼각형에 수직 방향을 계산함으로써 얼굴의 포즈를 계산한다. 계산된 얼굴의 포즈를 3D 얼굴 모델에 적용해 본 결과 본 논문에서 제안된 방법이 정확한 얼굴 포즈를 추정할 수 있음을 알 수 있었다.

Abstract

In this paper, we present an estimation method of a face pose by using two camera images. First, it finds corresponding facial feature points of eyebrow, eye and lip from two images. After that, it computes three dimensional location of the facial feature points by using the triangulation method of stereo vision techniques. Next, it makes a triangle by using the extracted facial feature points and computes the surface normal vector of the triangle. The surface normal of the triangle represents the direction of the face. We applied the computed face pose to display a 3D face model. The experimental results show that the proposed method extracts correct face pose.

▶ Keyword : Face Pose Estimation(얼굴 포즈 추정), Facial Feature Extraction(얼굴 특징 추출), Stereo Vision(스테레오 비전)

• 제1저자 : 소인미

• 접수일 : 2006.06.22, 심사일 : 2006.07.05, 심사완료일 : 2006.07.17

* 원광대학교 컴퓨터공학과 대학원 *** 원광대학교 전기전자및정보공학부 교수

※ 이 논문은 2006년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 헬스케어기술개발사업 지원을 받아 수행된 연구임.

I. 서론

컴퓨터 기술의 발전이 발전함에 따라 마우스나 키보드 같은 일반적인 입력장치에 의존하는 사용자 인터페이스를 탈피하여 보다 자연스럽게 편리한 사용자 인터페이스를 개발하려는 연구가 진행되고 있다. 많은 연구자들이 문자 인식, 음성 인식, 그리고 영상 인식 기술을 개발하고 이를 사용자 인터페이스에 활용하려는 연구를 진행하고 있다[1,2]. 그중에서도 영상 인식을 사용자 인터페이스에 이용하려는 연구가 근래에 들어 활발하게 진행되고 있다. 영상 인식을 이용한 사용자 인터페이스 개발에 있어서 사용자의 얼굴 포즈 추정은 중요한 기본 요소로 활용될 수 있다. 예를 들어, 사용자가 어떤 말을 하고 있는지를 입술의 움직임 정보로부터 인식하는 립리딩(lipreading)에서는 사용자가 어떤 방향을 보고 이야기 하고 있는지가 중요한 정보로 사용될 수 있다. 또한 제스처 인식을 이용한 사용자 인터페이스에서는 머리 제스처로부터 사용자의 의사가 긍정인지 부정인지를 인식할 수 있는데, 얼굴 포즈의 변화를 이용하여 사용자의 머리 제스처를 인식할 수 있다.

본 논문에서는 두 대의 카메라 영상으로부터 얼굴 포즈를 추정하는 방법을 제안한다. 제안된 방법에서는 한 사람의 얼굴 영상이 가까운 거리에서 두 대의 카메라에 의해 촬영되었다고 가정하고 입술의 양 끝 점이 보이는 데까지만 얼굴을 움직인다고 가정한다. 제안된 방법에서는 두 카메라 영상으로부터 눈과 눈썹의 안쪽 끝점 그리고 입술의 양 끝점을 찾은 다음에 그 점들의 3차원 좌표를 계산함으로써 얼굴 포즈를 추정한다.

본 논문의 II 장에서는 관련 연구를 설명하고 III 장에서는 얼굴 특징점 추출 방법에 대하여 설명한다. IV 장에서는 얼굴 포즈 추정 방법에 대해서 설명한다. 그리고 V 장에서는 실험 결과를 설명하고 VI 장에서는 결론과 앞으로의 연구 방향을 기술한다.

II. 관련 연구

기존의 얼굴 포즈 추정 방법은 패턴 분류에 의한 방법과 기하학 기반의 방법으로 분류될 수 있다. 패턴 분류에 의한 방법에서는 얼굴 포즈에 따라 얼굴의 모양이 달라지는 특성을 이용한다. 다양한 얼굴 이미지와 그에 대한 얼굴 포즈를 학습 데이터로 사용하여 학습시킨 다음에 새로운 얼굴 이미

지가 주어지면 학습된 정보를 활용하여 얼굴 포즈를 추정한다. 얼굴 포즈 추정을 위한 패턴 분류 방법으로는 주성분 분석(Principal Component Analysis)[3,4], 인공 신경망을 이용한 방법[5,6], SVM(Support Vector Machine)을 이용한 방법[7,8] 등이 제안되었다.

얼굴 모양은 사람에 따라 달라지고, 같은 사람의 경우에도 기분이나 조명과 같은 많은 주위 환경에 영향을 받으므로, 패턴 분류에 의한 방법에서는 여러 사람에 대하여 다양한 환경에서 촬영된 많은 양의 훈련 예제를 구축해야 한다는 문제를 가지고 있다. 또한 패턴 분류에 의한 방법에서 얼굴 포즈 추정의 정확도는 학습 데이터에서 얼굴 포즈를 몇 가지로 구분했는지에 영향을 받게 되는데, 얼굴 포즈를 세밀하게 구분하여 학습시키는 것은 아주 어려운 일이다. 따라서 대부분의 방법에서는 몇 가지 대표적인 얼굴 포즈에 대해 학습을 시키고 주어진 얼굴 이미지가 학습된 포즈에 정확히 일치 하지 않는 경우에는 보간법을 이용하여 얼굴 포즈를 추정한다. 이러한 이유로 인하여 패턴 분류에 의한 방법에서는 얼굴 포즈를 세밀하게 추정하는 데에는 한계가 있다.

기하학 기반의 방법에서는 얼굴의 구조에 대한 정보를 이용한다. 얼굴의 특징점을 이용한 방법[9,10]에서는 얼굴의 대표적인 특징점들을 추출한 다음에 이들의 기하학적 위치를 이용하여 얼굴 포즈를 추정한다. 또한 얼굴 특징점을 추출하는 데에도 얼굴 구성 요소들에 대한 기하학적 정보를 활용하였다. 3D 얼굴 모델을 이용한 방법[11,12]에서는 얼굴의 3차원 모델과 2차원 테스트 이미지 사이에 존재하는 얼굴 특징점들의 대응 관계를 이용하여 얼굴 포즈를 추정한다. 즉, 특징점들이 가장 잘 맞는 방향으로 3D 모델의 포즈를 변경해 나감으로써 얼굴 포즈를 추정한다. 기하학적 기반 방법은 구현이 용이할 뿐만 아니라 얼굴 특징점들의 위치가 정확한 경우에 얼굴 포즈의 정확도가 패턴 분류 기반 방법보다 훨씬 더 뛰어나다. 반면에 기하학적 기반 방법은 그 결과가 얼굴의 특징점 추출 결과에 민감하게 반응하고 특징점을 추출하기 어려울 정도로 얼굴을 많이 움직인 경우에는 사용하기 어렵다는 문제점을 가지고 있다.

기하학적 기반 방법들 중에는 스테레오 영상을 사용하는 방법도 제안되었다. 참고문헌 [13]에서는 스테레오 영상과 3D 모델을 이용한 방법이 제안되었다. 이 방법에서는 3D 모델을 입력 얼굴 영상에 맞추는 작업을 스테레오 비전의 에피폴라 제한(epipolar constraint) 특성을 이용하여 보다 정확하고 효과적으로 수행하는 알고리즘을 제안하고 있다. 그러나 이 방법에서는 얼굴 포즈의 추적을 시작할 때에

3D 모델을 추적 대상인 사람에 수작업을 통하여 맞추어야 하는 문제를 가지고 있다. 참고문헌 [14]에서는 캘리브레이션되지 않은 임의의 두 카메라로부터 입력된 영상으로부터 얼굴 포즈를 추정하는 방법을 제안하고 있다. 이 방법에서는 임의의 두 카메라 영상을 입력으로 받을 수 있는 장점이 있는 반면에, 얼굴의 특징 추출을 위해 수작업으로 얼굴 특징을 지정해주어서 학습을 시켜야 하는 단점을 가지고 있다. 참고문헌 [15]에서는 두 대의 카메라를 이미지 처리 보드에 연결해서 빠른 시간에 얼굴의 포즈를 추적하는 방법을 제안하고 있다. 이 방법에서는 얼굴의 특징을 추출하기 위해 눈, 코, 입 등의 영상을 입력 영상과 비교하는 방법을 사용하고 있다. 눈, 코, 입 모양 등에 대한 학습 과정을 필요로 한다. 본 논문에서는 영상의 명암도, 계곡 에너지 등 영상 자체의 속성만으로 얼굴 특징을 추출하고 스테레오 비전 기법을 이용하여 얼굴의 포즈를 계산하는 방법을 제안한다.

III. 얼굴 특징점 추출

3.1 얼굴 영역 검출

본 논문에서는 먼저 입력 영상으로부터 얼굴 영역을 검출한다. 얼굴 영역 검출을 위해서는 Viola에 의해 제안되고 Lienhart에 의해 개선된 다단계 분류기법[16, 17]을 이용하였다. 이 방법에서는 먼저 수백 장의 동일한 크기의 얼굴 영상과 얼굴이 아닌 임의의 영상을 사용하여 분류기를 학습시킨다. 분류기가 학습된 다음에는, 탐색 윈도우를 전체 입력 영상위에서 움직이면서 탐색 윈도우의 영상을 분류기에 전달하여 얼굴 영상인지 아닌지를 판별한다. 입력 영상에서 얼굴의 크기가 고정되어 있지 않고 변화할 수 있으므로, 이 분류기는 다양한 크기의 얼굴 검출을 효과적으로 수행할 수 있도록 분류기 자체가 계층적으로 설계되었다. 그림 1에는 차량 내부에서 두 대의 카메라로 촬영된 영상에 대한 얼굴 영역 검출 결과가 나타나 있다.

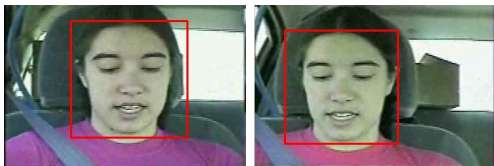


그림 1. 얼굴 영역 검출 결과
Fig 1. The result of face detection

3.2 얼굴 특징점 검출

얼굴 특징점으로는 그림 2와 같이 눈과 눈썹의 안쪽 끝점과 입술의 양쪽 끝점을 사용한다.



그림 2. 얼굴 특징점
Fig 2. Facial feature points

3.2.1 전처리

본 논문에서는 내부적으로 흑백 영상을 이용하므로 컬러 영상이 입력되면 먼저 흑백 영상으로 변환하고 히스토그램 평활화를 수행하여 영상의 명암 대비를 높인다.

3.2.2 얼굴 특징 에너지 계산

본 논문에서는 계곡 에너지, 명도 에너지, 경계선 에너지의 세 가지 에너지를 결합하여 사용함으로써 외부 환경 변화에 견고한 특징 추출 방법을 제안한다. 계곡 에너지는 그림 3과 같이 위와 아래는 밝고 가운데는 어두운 정도를 측정하는 에너지를 의미한다.

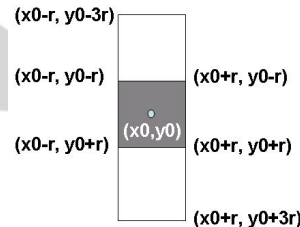


그림 3. 계곡 에너지 측정을 위한 패턴
Fig 3. The pattern for valley energy estimation

계곡 에너지는 픽셀을 중심으로 그림 3과 같은 형태의 사각형을 설정하고 위, 아래 사각형과 가운데 사각형의 명암도의 차이 값을 이용해서 구한다. 이때 변수 r 값을 일정 범위 내에서 변화시켜 가면서 명암도 차이가 최대가 되는 값을 선택한다. 이는 사람마다 얼굴 특징의 두께가 달라질 수 있기 때문이다. 픽셀 (x_0, y_0) 에 대한 계곡 에너지를 구하기 위해서는 먼저 식 (3.1)과 같이 가운데 사각형이 위와 아래의 사각형보다 어두운 정도를 나타내는 값 $V(x_0, y_0)$ 를 구한다.

여기에서 $p(x,y)$ 는 픽셀 (x,y) 의 명도값을 나타낸다. 기호 \max 는 변수 r 의 값이 l 부터 k 까지 변화할 때에 괄호안의 값의 최대값을 취하라는 의미이다. 그리고 l 과 k 의 값은 얼굴 영역의 크기에 따라 상대적으로 정해진다.

$$\begin{aligned}
 I(x_0, y_0) = & \max_{r=l \text{ to } k} \left\{ \left(\sum_{x=x_0-r}^{x_0+r} \sum_{y=y_0-3r}^{y_0+r} p(x,y) - \sum_{x=x_0-r}^{x_0+r} \sum_{y=y_0-r}^{y_0+r} p(x,y) \right) + \right. \\
 & \left(\sum_{x=x_0-r}^{x_0+r} \sum_{y=y_0+r}^{y_0+3r} p(x,y) - \sum_{x=x_0-r}^{x_0+r} \sum_{y=y_0-r}^{y_0+r} p(x,y) \right) - \\
 & \left. \left| \sum_{x=x_0-r}^{x_0+r} \sum_{y=y_0-3r}^{y_0+r} p(x,y) - \sum_{x=x_0-r}^{x_0+r} \sum_{y=y_0+r}^{y_0+3r} p(x,y) \right| \right\} \dots\dots\dots (3.1)
 \end{aligned}$$

전체 픽셀에 대해 $V(x_0, y_0)$ 를 구한 다음에는 계곡 에너지 $E_v(x_0, y_0)$ 를 식 (3.2)와 같이 구한다.

$$E_v(x_0, y_0) = \begin{cases} 0 & V(x_0, y_0) \leq 0 \\ \frac{V(x_0, y_0) \times 255}{Max_val} & V(x_0, y_0) > 0 \end{cases} \dots\dots\dots (3.2)$$

명도 에너지는 픽셀의 명도 값과 그 픽셀을 중심으로 한 5×5 크기의 영역의 평균값과의 차이를 이용해서 구한다. 이 값이 크면 클수록 주위 보다 더 어둡다는 것을 나타낸다.

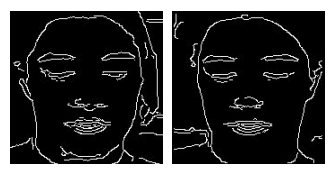
경계선 에너지는 Canny의 경계선 검출 알고리즘을 사용하였다. 이 알고리즘에서는 먼저 가우시안 필터를 이용하여 영상을 흐리게 한 다음에 1차 미분 연산자를 이용하여 픽셀 값의 변화율과 방향을 계산한다. 변화율이 가장 큰 값을 갖는 픽셀들이 연결되어 있는 부분이 경계선이므로, 픽셀 값의 변화 방향을 추적하여 경계선을 구한다. 그림 4에는 그림 1의 두 영상의 얼굴 영역에 대한 특징 에너지 계산 결과가 나타나 있다.



(a) 명도 에너지



(b) 계곡 에너지



(c) 경계선 에너지

그림 4. 얼굴 특징 에너지 계산 결과
Fig 4. The result image of facial feature energy computation

얼굴 특징 에너지를 계산한 다음에는 이들을 더해서 전체 에너지를 구한다. 전체 에너지를 구한 다음에는 눈과 눈썹 영역을 먼저 검출하고 그 다음에 입술 영역을 검출하는데, 영역 검출을 용이하도록 하기 위해 전체 에너지 이미지를 이진화하고 모폴로지 연산을 적용하였다. 그림 5에는 모폴로지 연산이 적용된 이진 영상이 나타나 있다.



(a) 전체 에너지



(b) 모폴로지 연산이 적용된 이진 영상

그림 5. 이진화와 모폴로지 연산의 결과 영상
Fig 5. The result image of binarization and morphology operation

3.2.3 얼굴 특징 영역 검출

그림 5에 나타나 있듯이 에너지 영상에서 눈, 눈썹, 입술 영역은 가로 방향의 흰색 사각형으로 나타난다. 따라서, 에너지 영상을 스캔하면서 가로 방향 사각형을 탐색함으로써 얼굴 특징 영역을 검출한다. 그림 6에는 얼굴 특징 영역의 검출 결과가 나타나 있다.

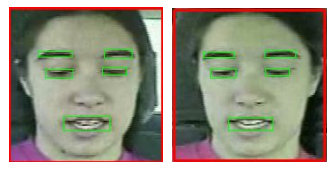


그림 6. 얼굴 특징 영역 검출 결과
Fig 6. The detection result of facial feature areas

3.2.4 얼굴 특징점 검출

얼굴 특징 영역을 검출한 다음에는 눈과 눈썹 영역의 안쪽 경계 부분과 입술 영역의 양 끝 부분을 대상으로 코너를 검출한다. 코너 검출을 위해서는 Harris 코너 검출 알고리즘을 사용하였다. 그림 7에는 검출된 얼굴 특징점들이 나타나 있다. Harris 코너 검출 알고리즘에서는 픽셀 밝기 값의 변화량에 대한 공분산 행렬을 구하고 그에 대한 고유값을 이용하여 코너를 검출한다.

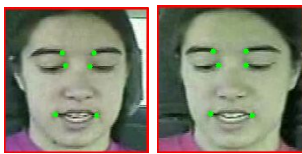


그림 7. 얼굴 특징점 검출 결과
Fig 7. The detection result of facial feature points

IV. 얼굴 포즈 추정

4.1 카메라 모델

카메라는 3차원의 점들을 2차원의 이미지 평면위로 투영시키는 장치이다. 3차원의 한 점 $V = (X, Y, Z)^T$ 가 2차원 이미지 평면의 한 점 $v = (x, y)^T$ 에 투영된다고 할 때 이들의 관계는 식 (4.1)과 같이 나타낼 수 있다. 여기에서 f_x 와 f_y 는 카메라의 초점 거리를 나타내고 (p_x, p_y) 는 주점(principal point)을 나타낸다. 주점이란 카메라의 중심을 통과하는 선이 이미지 평면과 수직으로 만나는 점을 의미한다.

$$x = f_x \frac{X}{Z} + p_x \quad y = f_y \frac{Y}{Z} + p_y \dots\dots\dots (4.1)$$

식 (4.1)에서는 카메라 좌표계와 세계 좌표계(world coordinate)가 일치하는 것을 가정하고 있다. 카메라가 움직여서 두 좌표계가 일치하지 않는 경우에, 카메라의 움직임은 회전과 평행 이동 두 가지의 조합으로 나타낼 수 있다. 카메라가 세계 좌표계에 대하여 R 만큼 회전하고 T 만큼 평행 이동한 경우에는, 3차원 좌표에 먼저 R^{-1} 와

T^{-1} 만큼 회전 변환과 이동 변환을 적용한 다음에 위의 식을 적용하면 2차원 이미지 평면에서의 좌표를 구할 수 있다. 종합하면 3차원 공간의 한 점은 일련의 변환, 즉 회전 변환, 평행 이동 변환, 카메라 모델에 의한 투영 변환에 의해 2차원 이미지 평면의 한 점으로 투영된다. 이들 변환은 하나의 카메라 행렬 P 로 표현될 수 있고 이를 이용하여 투영을 다음 식과 같이 표현할 수 있다.

$$v = PV$$

4.2 얼굴 특징점에 대한 3차원 좌표 복원

그림 8에 나타나 있듯이 두 대의 카메라 이미지가 주어지면, 서로 대응되는 점에 대하여 3차원 공간으로의 역투영을 수행하여 3차원 좌표를 구할 수 있다. 그림에서 C와 C'은 카메라의 렌즈의 중심을 나타내고 v와 v'은 두 이미지 평면에서의 3차원 공간의 점 V에 대한 투영을 나타낸다.

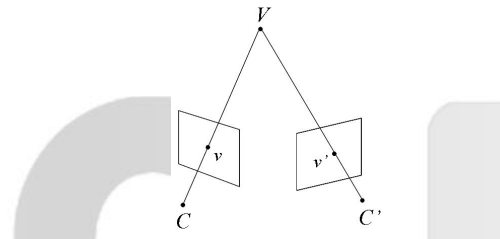


그림 8. 두 영상을 이용한 3차원 재구성
Fig 8. 3D reconstruction using two images

카메라 행렬이 각각 P 와 P' 이라 할 때 그림 8의 투영 관계는 $v = PV$, $v' = P'V$ 와 같이 나타낼 수 있다. v와 v'의 위치가 정확하면 카메라의 중심 C와 C'으로부터 v와 v'을 통과하는 두 선은 3차원 공간에서 만나게 된다. 두선이 3차원 공간에서 만나는 경우에는 방정식을 이용하여 V의 좌표를 간단하게 구할 수 있다. 그러나 카메라 이미지의 노이즈나 해상도의 한계로 인하여 v와 v'의 위치를 완벽하게 구할 수는 없고 결과적으로 두 선은 대부분의 경우에 공간상에서 만나지 않는다. 이런 경우에는 V의 좌표를 구하는 것은 복잡한 문제가 된다. 이 문제를 풀기 위해 여러 가지 방법이 제안되었는데 본 논문에서는 선형 최적화 방법을 사용한다. 이 방법에서는 $v = PV$, $v' = P'V$ 를 하나의 식 $AV = 0$ 의 형태로 결합한다. v와 v'이 정확하다면 $AV = 0$ 이 만족되고 연립 방정식을 풀어서 V의 값을 구할 수 있다. 그러나 v와 v'의 값에 노이즈가 있으면 $AV = 0$ 을 만족하는 V는

존재하지 않고 $\|AV\|$ 값을 최소화 하는 V 가 최적의 해가 된다. 이러한 선형 방정식의 최적화 해를 구하는 방법으로 여러 가지 방법이 있는데, 본 논문에서는 SVD(Singular Value Decomposition)를 사용하였다.

4.3 얼굴의 방향 계산

얼굴의 특징점들에 대한 3차원 좌표를 구한 다음에는 이들을 이용하여 얼굴의 방향을 계산한다. 본 논문에서는 삼각형에 대한 수직 벡터를 이용하여 얼굴의 방향을 계산한다. 삼각형의 꼭지점으로는 입술의 양 끝점, 그리고 두 눈썹 끝점의 중간점을 사용하였다. 삼각형이 그림 9와 같을 때에 이 삼각형에 대한 수직 벡터 n 을 구하기 위해서는 식 (4.2)와 같이 먼저 삼각형의 변을 나타내는 벡터 $E1$ 과 $E2$ 를 삼각형의 두 꼭지 점에 대한 벡터의 뽀셀연산으로 구한 다음에 벡터 $E1$ 과 $E2$ 의 외적(cross product)을 계산한다.

$$E1 = V2 - V1$$

$$E2 = V3 - V2 \dots\dots\dots (4.2)$$

$$n = E1 \times E2$$

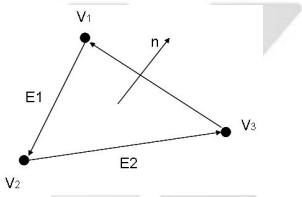


그림 9. 삼각형의 수직 벡터
Fig 9. The normal vector of a triangle

4.4 삼차원 얼굴 모델

앞에서 구한 얼굴 방향을 검증하기 위해 본 논문에서는 얼굴 방향 벡터를 3D 얼굴 모델에 적용해서 디스플레이 할 수 있도록 했다. 3D 얼굴 모델을 디스플레이하기 위해 OpenGL 라이브러리를 사용했다. 3차원 얼굴 모델은 그림 10과 같은 단순하게 얼굴만을 포함하도록 했고 기본적으로 z 축 방향을 보고 있는 것으로 모델링하였다. 따라서 카메라가 -z 축 방향을 향할 때에 얼굴의 방향이 정면이 된다. 주어진 영상에 대하여 얼굴의 방향을 구한 다음에는 그 방향이 정면 방향에서 얼마나 변환되었는지를 구할 수 있고 이 변환을 3D 얼굴 모델의 각 정점의 좌표에 적용하면 3D 얼굴 모델의 방향을 주어진 영상의 얼굴 방향과 일치시킬 수 있다. 그러나 3D 얼굴 모델의 각 정점 좌표를 변환하는

것 보다는 카메라의 위치를 변환하는 것이 더 간편하다. 즉, 얼굴이 움직인 방향과 반대 방향으로 카메라를 움직이면 얼굴을 움직인 것과 같은 효과를 얻을 수 있다.



그림 10. 3D 얼굴 모델
Fig. 10. 3D face model

V. 실험 결과

본 논문에서 제안된 방법은 윈도우 환경에서 Visual C++를 이용하여 구현되었으며, 달리는 자동차 내부에서 두 대의 카메라로 촬영된 영상을 이용하여 실험을 수행하였다. 그림 11에는 실험에 사용한 영상의 예가 나타나 있다. 이 영상은 미국 UIUC 대학에서 자동차 환경에서의 멀티 모달 음성 인식을 위해 구축한 데이터베이스[18]에 포함되어 있다.

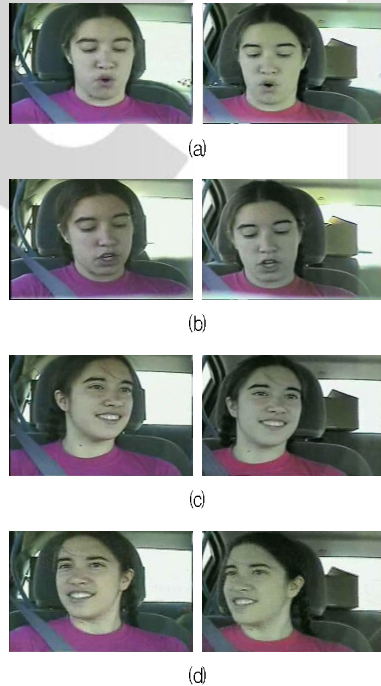


그림 11. 실험에 사용된 영상의 예
Fig 11. Image samples for the experiment

그림 12에는 본 논문에서 구현된 프로그램을 이용하여 그림 11의 영상에 대해 추출한 얼굴 포즈들이 순서대로 나타나 있다.

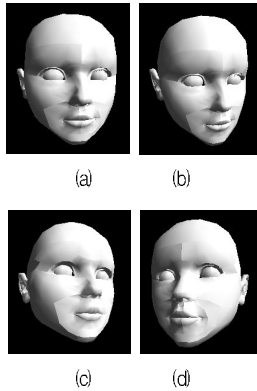


그림 12. 얼굴 포즈 추정 결과
Fig 12. Results of face pose estimation

실험에 사용된 두 카메라에 대한 카메라 행렬은 캘리브레이션 도구를 이용하여 구했으며 이들의 값은 다음과 같다.

$$P = \begin{bmatrix} 849.7765 & 0 & 152.2776 & 0 \\ 0 & 774.8115 & 123.2186 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P' = \begin{bmatrix} 823.7765 & 1.2807 & 256.9485 & -135041.0535 \\ -14.2964 & 769.7429 & 129.5156 & -944.04896 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

한 장면에 대한 얼굴 포즈 추정에는 0.26초의 시간이 소요되어서 초당 4 프레임의 영상을 처리할 수 있었다. 촬영된 동영상에 대해 실험을 한 결과 얼굴 포즈 추정 결과는 육안으로 판단하기에 올바른 포즈 추정 결과를 생성하였다.

본 논문에서 제안한 방법은 기존의 스테레오 영상을 이용한 방법에 비해서 얼굴 포즈 추출의 속도나 정확성 면에서 비슷한 성능을 보였다. 기존의 방법과 다른 점은 참고 문헌 [13, 14]에서는 얼굴 특징 추출을 추출하기 위해 별도의 수작업이 필요하였지만, 본 논문의 방법에서는 수작업이 필요 없이 자동으로 얼굴 특징을 추출하고 얼굴 포즈를 계산한다는 것이다.

본 논문에서 제안된 얼굴 포즈 추정 방법의 정확성은 얼굴의 특징점을 추출 결과에 영향을 받는다. 본 논문에서는 사람이 수작업으로 검출한 얼굴 특징점의 위치와 프로그램에 의해 자동으로 추출된 특징점의 위치를 비교함으로써 제안된 얼굴 특징 추출 방법의 성능을 테스트하였다. 비교를 위하여 참고문헌 [19]에 제안되어 있는 얼굴 특징 추출

방법을 구현하여 비교해 보았다. 얼굴 특징점의 위치를 검출한 다음에 사람이 수작업으로 검출한 위치와의 차이를 구하여 비교한 결과 본 논문에서 제안한 방법이 참고 문헌 [19]의 방법보다 50% 정도 오차가 적었다.

VI. 결론 및 앞으로의 연구 방향

본 논문에서는 두 대의 카메라 영상으로부터 얼굴의 특징점을 추출한 다음에 특징점의 3차원 좌표를 추출하고 삼각형의 수직 벡터 계산 방법을 이용하여 얼굴의 포즈를 추정하는 방법을 제안하였다. 제안된 방법에 의해 추정된 얼굴의 방향을 3D 얼굴 모델에 적용하여 디스플레이해본 결과 제안된 방법이 얼굴의 포즈를 효과적으로 추정할 수 있음을 알 수 있었다.

본 논문에서는 실시간 처리에 대한 고려를 하지 않았는데, 얼굴 포즈 추적 결과를 원격 회의나 인간과 컴퓨터의 상호 작용에 실제로 응용하기 위해서는 실시간 처리가 필수적이므로 이에 대한 연구를 수행할 예정이다. 그리고 본 논문에서는 일반적인 3D 얼굴 모델을 사용했는데, 입력 영상에 일치되는 3D 모델을 자동으로 또는 수동으로 생성함으로써 얼굴 포즈의 추적 결과를 보다 사실적으로 보여주는 연구를 수행해야 할 것이다.

참고문헌

- [1] 이희만, “스테레오 비전을 이용한 가상환경구현”, 한국컴퓨터정보학회논문지, 제 9권 3호, pp.79-85, 2004
- [2] 문형득, 구자영, “HMM을 이용한 지휘 동작의 인식”, 한국컴퓨터정보학회논문지, 제 9권 1호 pp. 25-30, 2004
- [3] S. Gong, S. McKenna and J.J. Collins, “An Investigation into Face Pose Distributions”, Proc. of ICAFG, pp. 265-270, 1996.
- [4] M. Motwani and Q. Ji, “3D face pose discrimination using wavelets,” Proc. of ICIP, pp. 1050-1053, 2001.
- [5] V. Krueger and G. Sommer, “Wavelet Networks for Face Processing”, Journal of the Optical Society of America, Vol. 19, No 6, pp. 1112-1119, 2002
- [6] L. Zhao, G. Pingali and I. Carlbon,

- "Real-time head orientation estimation using neural networks", Proc. of ICIP, pp. 297-300, 2002.
- [7] Jeffrey Huang, Xuhui Shao and Harry Wechsler, "Face Pose Discrimination Using Support Vector Machines (SVM)", ICPR, pp. 154-156, 1998.
- [8] J. Ng and S. Gong, "Composite Support Vector Machines for Detection of Faces across Views and Pose Estimation", Proc. of Image and Vision Computing, pp. 359-368, 2002.
- [9] Y. Hu, L. Chen, Y. Zhou, H. Zhang, "Estimating face pose by facial asymmetry and geometry", Proc. of ICAFGR, pp. 651 - 656, 2004
- [10] T. Horprasert, Y. Yacoob, and L.S. Davis, "Computing 3-d head orientation from a monocular image sequence," Proc. of ICAFGR, pp. 242-247, 1996.
- [11] P. Yao, G. Evans, and A. Calway, "Using affine correspondence to estimate 3-d facial pose," Proc. of ICIP, pp. 919-922., 2001,
- [12] G. Liang, H. Zha and H. Liu, "Affine Correspondence Based Head Pose Estimation for a Sequence of Images by Using a 3D Model", Prod. of ICAFGR..pp.632-637, 2004
- [13] R. Yang and Z. Zhang, "Model-based pose tracking with stereo vision", Proc. of ICAFGR, pp. 255-260, 2002
- [14] D. Gorodnichy, S. Malik and G. Roth, "Affordable 3D face tracking using projective vision", Proc. of ICVI, pp. 383-390, 2002.
- [15] Y. Matsumoto and A. Zelinsky, "An algorithm for real time stereo vision implementation of head pose and gaze direction implementation", Proc. of ICAFGR, pp. 499-505, 2000.
- [16] Paul Viola and Michael J. Jones. Rapid Object "Detection using a Boosted Cascade of Simple Features," Proc. of CVPR, 2001.
- [17] Rainer Lienhart and Jochen Maydt. "An Extended Set of Haar-like Features for Rapid Object Detection," ICIP, pp. 900-903, 2002.
- [18] B. Lee, M. Hasegawa-Johnson, C. Goudeseune, S. Kamdar, S. Borys, M. Liu and T. Huang, "AVICAR: Audio-Visual Speech Corpus in a Car Environment", International Conference on Spoken Language Processing, pp. 2489-2492, 2004
- [19] F. Domaika, J. Ahlberg, "Efficient active appearance model for real-time head and facial feature tracking", IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp.173 - 180, 2003

저자 소개



소인미

1994년 2월 : 원광대학교 컴퓨터공학과(공학사)
 2002년 8월 : 원광대학교 정보컴퓨터교육전공(교육학석사)
 2004년 3월 ~ 현재 : 원광대학교 컴퓨터공학과 박사과정
 관심분야 : 영상인식 컴퓨터 비전, 영상처리



강선경

2000년 2월 : 원광대학교 전기·전자공학부 졸업
 2004년 2월 : 원광대학교 정보·컴퓨터공학과 석사학위 취득
 2006년 2월 ~ 현재 : 원광대학교 컴퓨터 공학과 박사과정
 관심분야 : 휴먼 컴퓨터 인터페이스, 영상처리, 영상인식, 임베디드시스템



김 영 운

2003년 2월 : 원광대학교 전기전자
 및정보공학부(공학사)
 2005년 8월 : 원광대학교 컴퓨터
 공학과(공학 석사)
 2006년 2월 ~ 현재 : 원광대학교
 컴퓨터 공학과 박사과정
 1994-1999 영원한친구 개발팀장
 2000-현재 파라(PARA) 대표
 2003-현재 원광보건대학 정보컨텐
 츠과 겸임교수
 관심분야 : 컴퓨터그래픽스, 영상처
 리, EAI, XML



이 지 근

2000년 2월 : 원광대학교 컴퓨터공
 학과(공학사)
 2002년 2월 : 원광대학교 컴퓨터
 공학과(공학 석사)
 2002년 3월 ~ 현재 : 원광대학교
 컴퓨터 공학과 박사과정
 관심분야: 영상인식 영상인식, 컴퓨
 터 비전, 영상처리



정 성 태

1987년 2월 서울대학교 컴퓨터공학
 과(공학사)
 1989년 2월 서울대학교 컴퓨터공학
 과(공학석사)
 1994년 8월 서울대학교 컴퓨터공학
 과(공학박사)
 1995년 3월 현재 원광대학교 전기
 전자및정보공학부 교수
 관심분야 : 영상인식, 영상처리,
 컴퓨터 그래픽스

