

머리 움직임을 이용한 긍정/부정 의사 인식

문 병 선, 오 경 환
서강대학교 컴퓨터학과

Intention Recognition of Affirmation/Denial using Head Movement

Byoung-Sun Moon, Kyung-Whan Oh
Computer Science Dept. of Sogang University

요 약

본 논문은 고개를 심/하로 끄덕이거나 좌/우로 가로 저어서 긍정과 부정을 구별하기 위한 것이다. 다시 말해서, 마우스나 키보드 대신에 머리의 움직임을 사용해서 '예/아니오'를 인식한다. 본 논문에서는 정교화된 칼라 공간(chromatic color space)과 조도(illumination)를 이용하여 얼굴 영역을 찾고 분할하는 자동 얼굴 영역 찾기와 영상차의 위치 비교와 움직임 량을 이용하여 우선 순위를 갖는 단순한 방향성을 구별하는 자동 의사 인식의 두 단계로 구성되어 있다. 이러한 단순한 방향성의 조합으로 '예/아니오'를 구분한다.

1. 서론

사회가 발전하면 할수록 사람들은 더욱더 편리함과 편안함을 원하며 이에 따라서, 기술의 발전이 뒤따른다. 많은 분야에서 자동화가 이루어지고 사람의 의도를 알아서 주위의 환경들이 움직여 주기를 바란다. 머리 제스처는 장애가 있는 사람들이 컴퓨터를 제어할 수 있는 아주 매력적인 방법이다. 현재는 주로 키보드나 마우스를 주로 사용하고 있으며 이것은 손과 눈의 동시적인 협동, 움직임의 정확성과 반복성이 필요하다. 자유로운 움직임에 힘들거나 쉽지 않은 사람들에게는 문제가 된다. 이러한 문제의 대안으로 머리의 움직임을 파악하여 머리의 움직임을 이용한 머리 포인터 마우스나 어떠한 물음에 자신의 의사를 고개짓으로 표현할 수 있는 접근 방법을 생각해 볼 수 있다.

의사파악을 위해서는 두 가지 단계가 필요하다. 첫 번째 단계는 얼굴이 어디에 위치해 있는지 알 수 없는 영상에서 얼굴의 위치를 감지하고 찾아내는 자동 얼굴 영역 찾기(automatic human face location)이고 두 번째는 찾아낸 얼굴로부터 의사를 파악하는 자동 의사 파악(automatic human intention recognition) 단계이다. 성공적인 의사 파악 시스템을 위해서는 위의 두 가지 단계가 똑같이 강조되어야 하지만 현재 얼굴 위치 찾기의 경우 많은 연구가 진행되고 있는데 반해 자동 의사 파악의 경우는 연구가 거의 진행되고 있지 않다.

본 논문은 칼라 정보, 움직임 정보, 사전정보의 3가지 기술을 기반으로 한다. 제 2절에서는 전체적인 구조를 살펴보고,

제 3절에서는 실험 및 결과와 향후과제에 대해서 살펴보고 하겠다.

기존의 연구[5][6][7]들은 실시간 수행에 적용하기 어렵고 조명 밝기의 변화, 가림(occlusion)등의 영향에 민감하며, 특히 눈, 코, 입과 같은 얼굴 특징점으로 얼굴을 찾는 것은 많이 사용되고 있으나 시간이 변화할 때마다 이러한 특징점은 변화하기 때문에 정확한 눈의 움직임을 얻기 위한 시간과 연속적이고 정확한 움직임을 검출하기 어렵다. 가림(Occlusion)과 변형가능성(non-rigidity)의 경우는 큰 문제가 된다. 예를 들면 사람이 머리를 돌릴 때 얼굴 특징점이 가리거나 변형되기 때문이다. 많은 움직임 추정 알고리즘들은 고정(rigid) 물체에 대해서만 작동한다. 그러나 얼굴은 눈과 입 그리고 얼굴 근육 등이 변형 가능하기 때문에 고정(rigid) 물체로 여겨지지 않는다. 모델기반 추적(model-based tracking)과 변형 가능한 원형정합(deformable-template matching)이 이러한 특징점의 변화를 다루기 위해서 사용될 수 있다. 특징점의 다른 시점에 대응하는 각 원형을 가지고 하나의 특징점에 대해서 여러 개의 원형을 사용하는 것은 추적 성능을 향상시킬 수 있으나 계산시간이 많이 걸려 실시간 성능을 발휘할 수 없다[16]

2. 본론

전체적인 단계는 다음과 같고 이를 이해하기 쉽게 도식화한 그림 1이 다음에 있다.

- 단계 1. 영상 획득 및 칼라공간 변환
- 단계 2. 피부 칼라공간을 이용한 얼굴 위치 파악
- 단계 3. 얼굴 영역 추출 및 잡음 영역 제거
- 단계 4. 특징값 추출 및 영상비교 알고리즘
- 단계 5. 제스처 인식

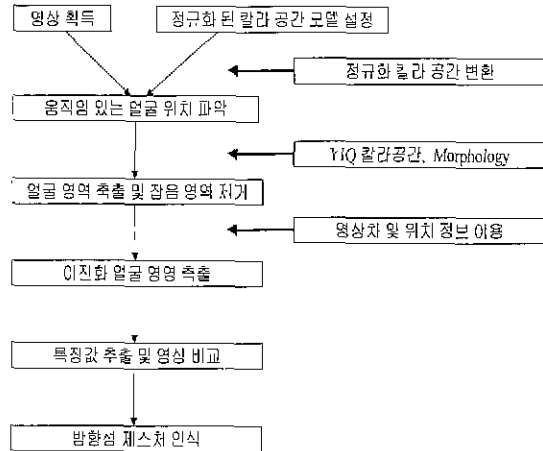


그림 1. 전체적인 시스템

단계1. 영상 획득 및 칼라공간 변환

영상 취득 과정은 컴퓨터가 사용자에게 물음을 한 후 어느 일정 시간 후부터 작동하도록 한다. CCD 카메라 및 비디오 캡코더 등의 장비를 이용하여 고개를 끄덕이거나 가로짓는 경우를 촬영한다. 영상에서 계속적인 움직임을 갖고며 조명과 배경의 변화에 계속적으로 적응해 갈 수 있게 하기 위하여 정규화 칼라 분포 공간을 이용한다. 즉, 정규화된 피부 분포가 정규 분포를 이루기 때문에 얼굴의 존재를 추적하기 위해서는 평균과 분산만을 계속해서 갱신한다.

단계2. 피부 칼라 공간을 이용한 얼굴 위치 파악

얼굴의 존재와 위치를 찾기 위해서 이미 만들어진 모델과 입력 영상을 정합(matching)한다. 입력 영상의 각 화소를 정규화 된 칼라 공간으로 변환하고 피부색 분포 모델과 비교한다. [3][4]

피부색은 일반 칼라 공간의 분포 범위보다 정규화된 칼라 공간에서 작은 범위로 나타나기 때문에 정합 과정이 빠르다. N개의 얼굴 영역만을 얻어낸 영상으로부터 관찰된 칼라 벡터 $\vec{C} = (r, g)$ 은 정규분포를 이루는 $p(\vec{C} | \text{피부색})$ 이 주어지므로 bayes rule을 이용하여 $p(\text{피부색} | \vec{C})$ 을 구한다.

$$P(\text{피부색} | \vec{C}) = \frac{P(\text{피부색}) P(\vec{C} | \text{피부색})}{P(\sim\text{피부색})P(\vec{C} | \sim\text{피부색}) + P(\text{피부색})P(\vec{C} | \text{피부색})} \quad (1)$$

단계3. 얼굴 영역 추출 및 잡음 영역 제거

정규화 칼라 분포를 통해 얻은 영상에는 여러 가지 크고 작은 잡음들이 존재한다 그러므로 명확하고 깨끗하게 얼굴영역을 얻어내기 힘들다 여기서 얼굴 사각형 이외의 영역은 고려할 필요가 없기 때문에 얼굴 사각형을 이루는 M X N 마스크를 적용한 convolution 연산을 수행해서 배경에 존재하는 얼굴 위나 옆의 제거되지 않고 남아 있는 어느 일정 크기 이하의 작은 잡음들을 제거한 후, 조도(illumination)값으로 [I] - RGB를 칼라공간의 값을 YIQ 또는 HSV로의 변환하였다. - 변환하고 임계치를 설정하여 배경이 되는 부분은 0으로 나타내고 얼굴 부분을 255값을 갖는 그레이 값으로 표현한다. 이러한 과정을 거치면 얼굴 영역을 얻어낼 수 있으나 얼굴 경계나 주위에 존재하는 뭉쳐져 있는 작은 크기의 잡음과 얼굴내의 잡음이 존재한다. 이러한 잡음은 모폴로지(부식(erosion)-팽창(dilation) -팽창(dilation)-부식(erosion)의 단계)를 사용하여 제거하였다.

$$(A \ominus B) \oplus B = A \cdot B \quad (2)$$

$$[(A \cdot B) \oplus B] \ominus B = (A \cdot B) \cdot B$$

단계4. 특징값 추출 및 영상 비교 (영상비교 알고리즘)

고개를 좌측이나 우측으로 가로 저어졌는지를 알기 위해서는 다음과 같은 순서를 따른다. 우선 좌우를 대칭적으로 가를 수 있는 대략적인 대칭축을 구한다. 얼굴 영역 내에서 세로 최대함을 갖는 축과 얼굴 영역의 가로축을 구적법으로 구한 후 이 둘의 평균을 구해서 좌/우측을 가를 수 있는 대략적인 대칭축을 얻는다. 전 영상과 후 영상의 차를 얻어서 대칭축을 기준으로 좌측이나 우측으로 움직임을 나타내는 화소들의 합이 많은 곳으로 움직였다고 나타낸다 다음으로 상하를 가르기 위해서는 일정 폭의 크기(대체로 이마 부분에 해당)로 제한된 가로축의 전/후 영상간의 위치 변화와 코 부분에 해당하는 얼굴에서 가장 큰 폭을 갖는 가로 최대합 축의 전/후 영상의 위치 변화를 파악하여 상/하로의 얼굴의 끄딤을 알아낸다. 최종적인 상/하/좌/우의 판별은 이것들의 서로 반복적이고 의존적인 조건을 가진 지식기반(knowledge-based)에 따른 알고리즘에 의해서 우선순위를 갖는 결과가 각각 나오고 이것들의 조합에 의해서 '예/아니오'가 결정된다 자세한 파라미터 및 알고리즘은 지면판계상 생략한다.

단계5. 제스처 인식 (머리 제스처 판단 알고리즘)

- ① 세 개의 (I_1, I_2, I_3) 영상을 칼라 공간과 움직임을 이용한 머리 제스처 알고리즘에 (I_1, I_2)와 (I_2, I_3)의 쌍으로 넣는다
- ② 각각의 쌍으로부터 나온 우선순위(priority)를 비교하여 결과를 얻는다.

3. 실험 결론 및 향후과제

· 단순한 방향성 제스처

이 제스처는 주요 축에 따른 머리의 움직임을 나타낸다. 머리에 대해서 4개(위, 아래, 좌, 우)의 제스처가 있다. '위'는 머리를 뒤로 젖혀서 나타나고 '아래'는 가슴을 향해 머리를 내림으로써 나타난다. '좌'는 머리를 둥글게 돌리는 것이라고 하기 보다 좌로 돌리는 것을 말한다 '우'는 머리를 우측으로 돌리는 것을 나타낸다.

· 진동하는 방향성 제스처

'예/아니오'의 두 가지 제스처가 있다. 이것은 단순한 방향성 제스처의 결합으로부터 나온다. 마우스의 경우 정해진 시간간격으로 마우스 버튼이 '내려감-올라감-내려감-올라감'의 반복적인 움직임을 감지하여 더블클릭이 일어났는지 안 일어났는지 알 수 있다. 이와 같은 방법으로 '아니오'는 머리를 좌우로 흔드는 몇 개의 '좌', '우' 제스처로 이루어진다.

앞으로의 연구에서는 현재의 시스템을 실시간으로 추적하는 부분을 보강해서 재귀적이면서 연속적으로 방향성을 추적할 수 있도록 머리의 움직임과 카메라의 움직임을 조정해주는 것이 필요로 하다. 또한 방향성 판별과 의사 판별을 위한 비교 알고리즘에 있어서도 지금과 같이 비교적 불안정한 상대적인 비교 방법이 아닌 앞의 결과가 뒤의 결과에 영향을 미치는

표 1 실험 결과(%)

정면 기준 머리 크기 (°)	좌/우			상/하		
	< 5	< 10	< 15	< 5	< 15	
비디오 캠코더						
좌측	93	94	96	상	87	93
우측	94	96	96	하	84	90
좌/우측	100	100	100	상/하	88	95
CCD 카메라						
좌측	85	90	95	상	77	88
우측	84	91	94	하	77	86
좌/우측	86	92	95	상/하	85	89

계속적인 적응이 가능한 좀 더 좋은 방법과 특징점을 명확하게 선택해 주어서 특징점의 이동 상태 등을 관찰 해 상대적인

방향성을 판별해 줄 수 있으면 더욱 좋은 결과를 얻을 수도 있을 것이다

본 논문은 기존의 마우스 역할을 대신해서 좀 더 사용자에게 편리함을 제공해줄 수 있으며 앞으로 응용 가능성도 크다고 생각된다.

참고 문헌

- [1] Y. Dai and Y. Nakano, "Face-Texture Model based on SGLD and Its Application in Face Detection in a Color Scene," Pattern Recognition, Vol. 29, No. 6, pp 1007-1017, 1996.
- [2] H. B. Kim and R. H. Park, "Extracting Spatial Arrangement of Structural Textures using Projection Information", Pattern Recognition Society, Vol 25, No 3, pp 237-245, 1992
- [3] H.A. Rowley, S. Baluja, and T. Kanade, "Human Face Detection in Visual Scenes," Technical Report CMU-CS-95-158, CS department, CMU, 1995.
- [4] J. Yang and A. Waibel, "Tracking Human Faces in Real Time," CMU CS Technical Report, CMU-CS-95-210, November, 1995.
- [5] C. H. Lee, J. S. Kim and K. H. Park. "Automatic Human Face Location in a Complex Background using Motion and Color Information, Pattern Recognition," Pattern Recognition, Vol 29, No 11, pp 1877-1889, 1996.
- [6] 최현일, 김수환, 최정일, 이필규, "멀티미디어 휴먼 인터페이스를 위한 헤드 제스처 인식," '98 봄 학술 발표논문집(B) 제25권 1호, pp. 639-641, 1998
- [7] A. D. Wilson and A. F. Bobick, "Configuration States for the Representation and Recognition of Gesture," Proceedings International Workshop on Automatic Face and Gesture-Recognition, Zurich, 1995