

# 얼굴의 포즈 정보를 이용한 헤드 제스처 인식에 관한 연구

김정연<sup>†</sup>, 박형철<sup>\*\*</sup>, 전병환<sup>†</sup>  
<sup>†</sup> 공주대학교 전자계산학과, <sup>\*\*</sup>(주)에코솔루션

e-mail : bhjun@knu.kongju.ac.kr

## The Recognition of Head Gestures using Face Pose Information

Jeong Yeon Kim<sup>†</sup>, Hyung Chul Park<sup>\*\*</sup>, Byung Hwan Jun<sup>†</sup>  
<sup>†</sup> Dept. of Computer Science, Kongju National University,  
<sup>\*\*</sup>ECO Solutions Co., Ltd

### 요약

본 논문에서는 얼굴의 9가지 상태를 인식하고 이를 상태 시퀀스로 생성한 후, 오토마타 기법을 적용하여 13가지(준비, 상측, 하측, 좌측, 우측, 전진, 후퇴, 좌 윙크, 우 윙크, 좌 더블 윙크, 우 더블 윙크, 긍정, 부정) 헤드 제스처를 인식하는 방법을 제안한다. 얼굴 영역을 추출하는 방법에서는 최적의 얼굴색 정보와 적응적 움직임 정보를 이용하여 얼굴 영역을 추출한다. 눈의 후보 영역을 추출하는 방법에서는 소벨 연산자와 투영 기법을 이용한다. 이 때 눈의 후보들을 제거하기 위하여 눈의 기하학적 정보와 눈은 쌍으로 존재한다는 정보를 이용한다. 얼굴의 상태를 인식하기 위해서는 계층적인 특징분석 방법을 사용한다. 13가지 헤드 제스처는 얼굴 상태 인식의 처리에서 생성된 상태 시퀀스를 이용한 오토마타 기법에 의해 인식된다. 실험 결과, 93.3%의 헤드제스처 인식률을 얻었다.

### 1. 서론

컴퓨터 관련 학문 및 기술의 발전으로 인하여 기존의 한정된 입력 장치에 의존하는 사용자 인터페이스에 비해 보다 자연스럽고 편리한 사용자 인터페이스에 대한 요구가 증가하고 있다. 이러한 요구를 충족시키기 위해 많은 연구자들이 문자 인식, 음성 인식, 그리고 영상 인식 분야에 대해 연구를 진행하고 있다. 특히 영상 인식 분야는 화상 통제 시스템이나 지문 인식, 그리고 헤드 제스처 인식(head gesture recognition) 등과 같이 최근에 가장 활발히 연구되고 있는 분야이기도 하다. 이 중 헤드 제스처 인식은 사용자와 컴퓨터간의 의사 소통을 기계적인 방식(키보드, 마우스 등)에 의존하지 않고, 얼굴의 방향이나 고개의 움직임 또는 눈의 깜빡임 등과 같은

자연스러운 정보를 이용하여 컴퓨터 사용자로 하여금 흥미를 유발하게 하고 제한된 작업 환경을 탈피하도록 하여 보다 자연스러운 컴퓨터 사용 환경을 제공해 주는 것이 가능하다.

기존의 제스처 인식에 관련된 연구는 주로 긍정, 부정, 중립의 3가지 제스처에 대해 연구가 진행되어왔다. Morimoto 등[1]은 HMM을 이용하여 YES, NO, MAYBE, HELLO의 4가지 제스처를 정의하고 구현하였다. 문병선 등[2]은 영상차의 위치 비교 및 움직임 정보를 이용한 단순 방향성을 구별하여 긍정, 부정의 제스처를 인식하였다. 최현일 등[3]은 모자이크와 BP(Back Propagation) 신경망을 이용하여 얼굴 영역을 추출한 후, 세그멘테이션과 이진화를 이용하여 눈 영역을 추출하여 이 값을 HMM의 입력 값으로 이용하여 헤드제스처를 인식하였다. 이와는 다르게 Xu 등[4]은 적절한 임계값을 바탕으로 그레이(gray) 영상에서 모폴로지 오픈링(morphology opening) 연산자를 이용하여 얼굴 영역을 추출한 후, 이진화 및 투영기법에 의해 두 눈동자와 입을 추출하여 다양한

※ 본 논문은 한국과학재단 지정 지역협력 연구센터인 공주대학교 자원재활용 신소재 연구센터의 연구비 지원에 의해 연구되었음.

헤드 포즈(head pose)를 추출하였다. 그러나 이 방법은 배경이 단순한 그레이 영상에서 수행하였기 때문에 배경이 복잡한 컬러 동영상에서는 헤드 포즈 인식에 많은 문제점을 가지고 있다.

본 논문에서는 최적의 얼굴색 정보를 이용하여 얼굴 영역을 추출하고, 추출된 얼굴 영역에서 소벨 연산자 및 프로젝션 기법, 그리고 눈 영역에서의 화소값들의 차이 정보를 사용하여 눈 영역을 추출한다. 추출된 얼굴 영역과 눈 영역을 이용하여 얼굴의 방향, 이동, 그리고 눈의 뜨고 감은 상태 정보를 인식한다. 그러다음 생성되는 상태 시퀀스에 대해 오토마타 기법을 이용하여 헤드제스처를 인식한다.

그림 1은 본 논문에서 제시하는 방법의 전체 흐름도이다.

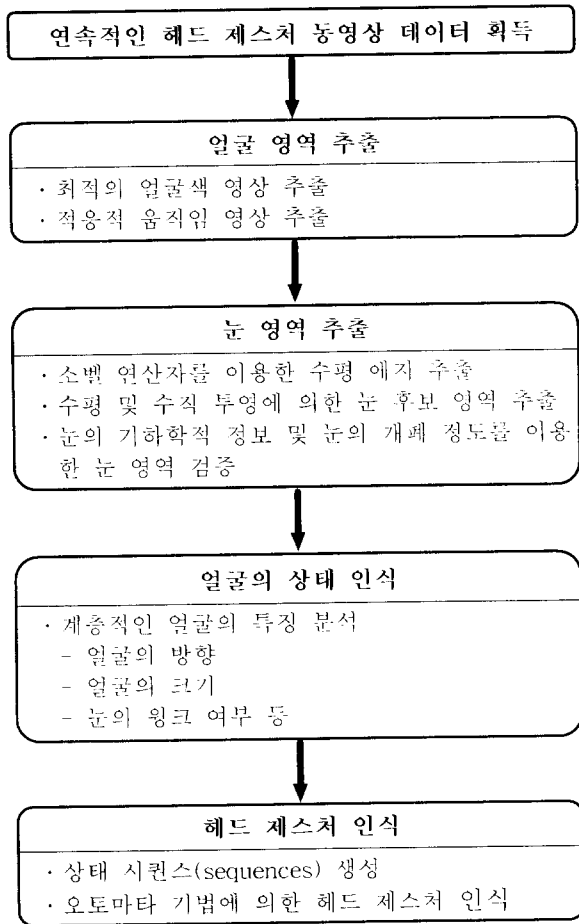


그림 1. 전체적인 인식 과정 흐름도

## 2. 얼굴 및 눈 영역 추출

### 2.1 얼굴 영역 추출

최근 동영상에서 얼굴 영역을 추출하는 방법으로 색상 정보 및 움직임 차이 정보를 많이 이용하고 있다[5]. 이러한 방법들은 처리

시간이 빠르고 복잡한 배경에서도 비교적 얼굴 영역을 잘 추출한다는 장점이 있다. 본 논문에서는 얼굴 영역을 추출하기 위해 최적의 얼굴색 정보와 움직임 정보를 이용한다[6]. 최적의 얼굴색 정보를 추정하는데는 YIQ 컬러 모델의 I 성분을 이용하였고, 움직임 정보의 경우에는 두 영상( $t$ ,  $t+1$ )간 일정한 차 정보가 발생하지 않으면 영상  $t$ 와  $t+1$ 의 다음 영상( $t+2$ )을 이용하는 방식을 통해 움직임 정보를 이용하였다. 그림 2는 최적의 얼굴색 정보를 이용한 영상과 차영상, 그리고 얼굴 영역이 추출된 결과를 나타낸다.

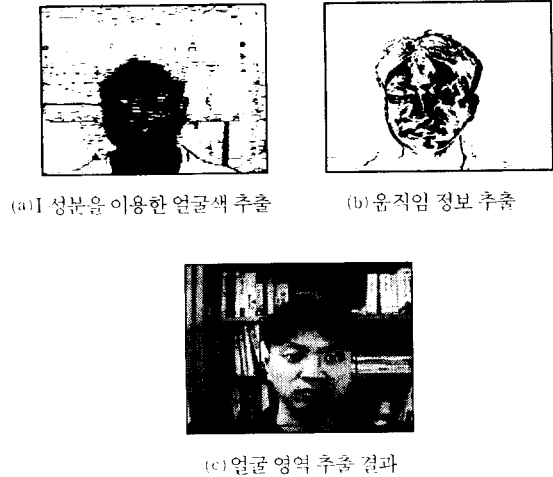


그림 2. 최적의 얼굴색과 움직임 정보를 이용한 얼굴 영역 추출 결과

### 2.2 눈 영역 추출

2.1절에서 성공적으로 얼굴 영역을 추출하면 그림 3과 같은 과정에 의해 눈 영역을 추출한다.

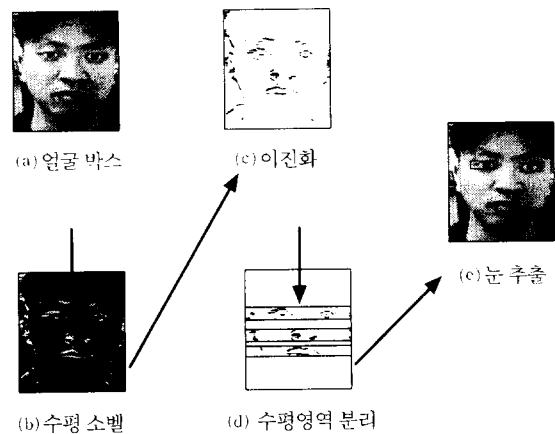


그림 3. 눈 영역 추출 과정 및 결과

먼저, 수평 소벨 연산자를 이용하여 얼굴 영역내의 수평 에지 성분을 추출하여 이진화한다. 이진화된 얼굴 영역에 대해 수평방향

으로 프로젝션을 수행한 후, 3개의 최고 마루를 기준으로 가장 근접한 최소 골을 찾아내어 영역 분리를 시도한다. 이렇게 분리된 영역들을 대상으로 수직방향으로 프로젝션을 수행하여 수평 영역 분리시 사용된 방법과 동일한 방법으로 눈 후보 영역들을 추출한다. 추출된 후보 영역들에 대해서 실제 눈 영역을 검증하기 위해 눈의 기하학적 위치 정보와 눈의 감은 정도를 분석한다. 눈의 감은 정도는 눈 영역에서 흰자위와 검은자위의 평균 명암도 차이를 이용한다.

### 3. 얼굴의 상태 인식

#### 3.1 얼굴의 9가지 상태 정의

본 논문에서 정의하는 얼굴의 9가지 상태는 그림 4와 같다.[7]



그림 4. 얼굴의 9가지 상태에 대한 예

#### 3.2 상태 인식에 사용되는 계층적인 특징

얼굴의 상태를 구분하는 특징은 표 1과 같다.

표 1. 얼굴의 9상태를 구분하기 위한 특징

구분 특징	설 명
특징 A (전후구분)	S: 얼굴 박스의 크기(변적) S <sub>G</sub> : 최근 정상 상태의 얼굴 박스 크기 S <sub>C</sub> : 현재 상태의 얼굴 박스 크기
	최근 정면 상태의 얼굴 박스 크기에 대한 현재 상태의 얼굴 박스 크기의 비율 $R_S = \frac{S_C}{S_G}$
특징 B (좌우구분)	L: 얼굴 박스의 가장자리까지의 거리 L <sub>L</sub> : 얼굴 박스의 좌측 가장자리까지의 거리 L <sub>R</sub> : 얼굴 박스의 우측 가장자리까지의 거리
	양 눈 사이의 중심에서 얼굴 박스의 좌측 가장자리까지의 거리에 대한 우측 가장자리까지의 거리 비율 $R_L = \frac{L_R}{L_L}$
특징 C (상하구분)	W: 차영상 얼굴 박스의 상단 끝까지의 거리 W <sub>G</sub> : 최근 정면 상태의 경우 W <sub>C</sub> : 현재 상태의 경우
	양 눈 사이의 중심에서 차영상 박스의 상단 끝까지의 거리 비율 $R_W = \frac{W_C}{W_G}$
특징 D (깜빡임구분)	D: 눈 영역에서 평균 화소값 상위 10%와 하위 10%의 차이 D <sub>L</sub> : 왼쪽눈의 경우 D <sub>R</sub> : 오른쪽눈의 경우
	눈 영역에서 화소값 상위 10%와 하위 10%의 차이를 구하고, 좌측 눈과 우측눈의 차이값 $D_D = D_L - D_R$
특징 E (방향구분)	특징 B와 특징 C를 이용한 상태 인식에 사용 최근에 정의된 정면 상태의 눈 위치와 현재 상태의 눈 위치 비교

그림 5는 계층적인 특징을 사용하여 얼굴 상태를 인식하는 과정을 나타낸 것이다.

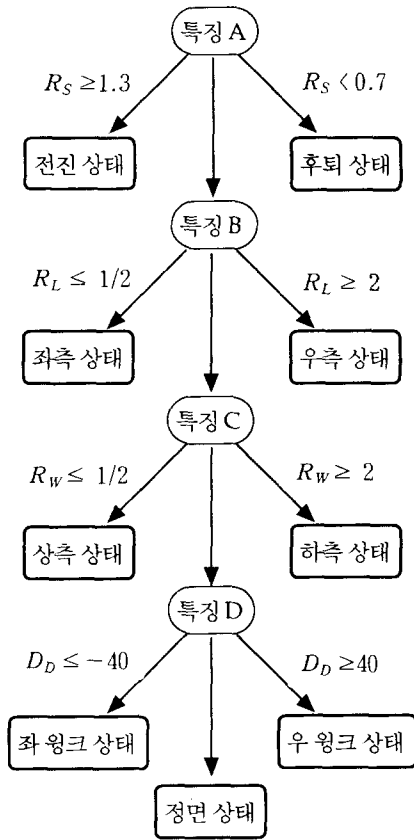


그림 5. 계층적인 상태 인식 특징을 이용한 상태 인식도

## 4. 헤드 제스처 인식

### 4.1 헤드제스처의 정의

본 논문에서 사용하는 헤드 제스처(head gesture)는 준비(Geze), 상측(Upward), 하측(Downward), 좌측(Leftward), 우측(Rightward), 전진(Forward), 후퇴(Backward)의 포즈 제스처와 좌 윙크(Left Wink), 우 윙크(Right Wink), 좌 더블 윙크(Left Double Wink), 우 더블 윙크(Right Double Wink), 긍정(Yes), 부정(No)의 행위 제스처로 총 13가지 헤드 제스처이다.

### 4.2 헤드제스처 인식방법

앞에서 정의한 헤드 제스처(head gesture)를 인식하기 위하여 본 논문에서는 오토마타 기법(automata method)에 의한 헤드 제스처 인식 방법을 제안한다. 3장에서 정의한 얼굴의 상태를 헤드 제스처 인식을 위한 상태 시퀀스(sequences)로 생성하고, 생성된 시퀀스들을 헤드 제스처로 인식하기 위해 표 2와 같은 오토마타 기법을 적용한다.

표 2. 헤드 제스처 인식에 사용되는 오토마타 기법

헤드 제스처	오토마타 기법
준비	$G \leftarrow g^e \mid Gg$
상측 이동	$U \leftarrow G u^e \mid Uu$
하측 이동	$D \leftarrow G d^e \mid Dd$
좌측 이동	$L \leftarrow G l^e \mid Ll$
우측 이동	$R \leftarrow G r^e \mid Rr$
전진	$F \leftarrow G f^e \mid Ff$
후퇴	$B \leftarrow G b^e \mid Bb$
좌 윙크	$W_L \leftarrow G w_L^+ \mid g$
우 윙크	$W_R \leftarrow G w_R^+ \mid g$
좌 더블 윙크	$W_{LD} \leftarrow W_L g^* w_L^+ \mid g$
우 더블 윙크	$W_{RD} \leftarrow W_R g^* w_R^+ \mid g$
긍정	$Y \leftarrow G d^+ g^+ u^+ g^+ d^+ g^+ u^+ \mid g$
부정	$N \leftarrow G l^+ g^+ r^+ g^+ l^+ g^+ r^+ \mid g$

표 3. 오토마타 기법에 사용되는 첨자

첨자	정의
$e$	3회 생성
$*$	0회 이상 생성, 최대 5회까지 생성
$+$	1회 이상 생성, 최대 5회까지 생성

인식에 사용되는 상태 시퀀스에 의하면 준비 제스처가 계속해서 생성될 수 있는데, 이는 각 제스처를 명확히 구별하기 위해 헤드 제스처 사이마다 준비 제스처가 오도록 동영상 데이터 입력시에 조절했기 때문이다. 그러나 본 논문에서는 준비 제스처를 처음 시작시에만 인정하여 중간 단계에서 생성되는 준비 제스처는 무시하는 방법으로 총 13종류의 제스처에 대해 인식 실험을 실시하도록 한다.

## 5. 실험

### 5.1 실험 환경

제안한 방법의 평가를 위해, Pentium MMX 200MHz PC에 CCD 카메라를 이용하여 320×240 크기의 동영상 데이터를 획득하였고, Visual C++ 6.0으로 알고리즘을 구현하였다. 실험에 사용된 동영상 데이터는 뒷배경이 복잡하고 컴퓨터 화면을 주시하는 얼굴 영상으로 총 8명의 데이터로 구성되어있다. 동영상 데이터는 정면, 좌측, 우측, 상측, 하측의 주시와 전진과 후퇴의 이동, 그리고 좌 윙크와 우 윙크의 순서로 연속적인 제스처를 취하여 획득하였다.

### 5.2 실험 결과 및 분석

본 논문에서 제안한 방법에 의해 실험된 결과가 표 4에 나타나 있다.

표4. 헤드 제스처 인식 결과  
(인식된 제스처 수/총 제스처 수)

구 분		헤드 제스처 인식률		
헤드 제스처	포즈 제스처	준비	100% (8/8)	100% (56/56)
		상측	100% (8/8)	
		하측	100% (8/8)	
		좌측	100% (8/8)	
		우측	100% (8/8)	
		전진	100% (8/8)	
		후퇴	100% (8/8)	
	행위 제스처	좌 윙크	87.5% (7/8)	85.4% (41/48)
		우 윙크	87.5% (7/8)	
		좌 더블 윙크	75.0% (6/8)	
		우 더블 윙크	87.5% (7/8)	
		궁경	87.5% (7/8)	
		부정	87.5% (7/8)	
		93.3% (97/104)		

실험결과, 포즈 제스처에서는 100%의 인식률을 얻었다. 포즈 제스처의 경우 상태 인식 과정에서 실패한 상태가 발생하더라도 오토마타 기법에 의해 제스처를 인식하는데 문제점이 발생하지 않는다. 한편 행위 제스처의 경우에는 실험용 동영상 데이터를 제공한 사람들의 신체적인 특성으로 인해 인식률이 하락한 경우가 발생하였다. 즉 윙크 제스처의 경우 사람에 따라 불완전한 윙크 동작을 취하는 경우로 인해 제스처 인식률의 하락을 가지고 왔다. 이에 따라

행위 제스처는 85.4%의 제스처 인식 성공률을 나타냈다. 결과적으로 본 논문에서 제안한 헤드 제스처 인식 방법은 전체적으로 93.3%의 헤드 제스처 인식 성공률을 나타냈다.

## 6. 결론

본 논문에서는 헤드 제스처 인터페이스(head gesture interface)를 구현하기 위해 헤드 포즈(head pose) 및 윙크(wink) 정보에 의한 얼굴의 9가지 상태를 인식하고, 인식된 상태들을 상태 시퀀스(sequences)로 생성하여 오토마타 기법(automata method)에 의한 13가지의 헤드 제스처를 인식하는 방법을 제안하였다.

헤드 제스처를 인식하기 위해서는 얼굴 및 눈 영역 추출이 선행되어야 한다. 얼굴 영역 추출에서는 HSI 컬러 모델의 H(hue) 성분과 YIQ 컬러 모델의 I(in-phase) 성분에 대해 최적의 얼굴색 정보를 이용한 비교 실험을 통하여, 얼굴 영역 추출에 보다 효과적인 I 성분의 최적 얼굴색 정보(구간: -22 ~ -2)와 적응적 움직임 정보를 이용하는 방법을 사용하였다. 눈 영역 추출 방법에서는 처리시간을 단축하기 위해 얼굴 박스 영역 안에서 소벨(sobel) 연산자를 이용하여 이진화된 수평 성분의 에지(edge)를 추출하고 수평 및 수직 투영(projection)을 통해 눈 후보 영역을 추출하였고 추출된 눈 후보 영역에 대해 눈의 기하학적 정보와 개폐 정도를 이용하는 눈 검증 방법을 사용하였다. 추출된 얼굴 및 눈 영역을 이용하여 얼굴의 9가지 상태를 구분하는 방법으로 얼굴의 크기 정보를 이용한 전후 상태 구분, 얼굴의 방향 정보를 이용한 상하좌우 구분, 그리고 눈의 윙크 정보를 이용한 좌우 윙크 구분 등의 계층적인 주 특징과 얼굴의 방향을 구분하는데 보조적으로 사용하는 눈 위치 이동 특징을 사용하였다. 인식된 얼굴의 상태를 상태 시퀀스로 생성하여 오토마타 기법에 의해 13종류의 헤드 제스처를 인식한 결과, 104개의 헤드 제스처에 대해 97개인 93.3%의 인식률을 나타내었다.

향후 연구 과제로는, 눈과 눈썹 영역이 결합되어 눈 추출에 실패하는 경우를 해결하기 위한 보다 정교한 눈 영역 분할 알고리즘의 개발과, 머리카락이 눈을 가리는 경우에 대한 해결책을 모색해야 한다. 그리고 본 논문에서 실험하지 않은 안경을 착용한 사람에 대해서도 헤드 제스처 인식의 성능을 평가해야 한다. 또한 헤드 제스처 인식을 위한 동영상 데이터가 정형화된 순서로 되어 있기 때문에 비정형화된 순서의 데이터에 대한 실험 및 평가를 통해 자유로운 헤드 제스처가 가능하도록 하는 연구도 필요하다. 마지막으로 실시간 처리에 적합하도록 헤드 제스처 인터페이스에 적용하는 연구가 필요하다.

## 참고문헌

- [1] C. Morimoto, Y. Yacoob and L. Davis, "Recognition of Head Gestures Using Hidden Markov Models", IEEE Proc. of Int'l Conf. on Pattern Recognition, pp.461-465, 1996.
- [2] 문병선, 오경환, "머리 움직임을 이용한 긍정/부정 의사 인식", 한국정보과학회 추계학술발표논문집, 제25권, 제 2호, pp.538-540, 1998.
- [3] 최현일, 김수환, 이필규, "시각사용자 인터페이스를 위한 Multi-classifier 기반의 헤드 제스처 인식", 한국정보처리학회 추계학술발표논문집, 제5권, 제2호, pp.1295-1298, 1998.
- [4] M. Xu and T. Akatsuka, "Detecting Head Pose from Stereo Image Sequence for Active Face Recognition", Proc. of Int'l Workshop on Automatic Face- and Gesture-Recognition, pp.82-87, 1998.
- [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, Vol. 29, No. 11, pp.1877-1889, 1996.
- [6] 박형철, 전병환, "동영상에서 최적의 얼굴색 정보와 움직임 정보에 기반한 얼굴 영역 추출", 한국정보과학회 논문지:소프트웨어 및 응용, 제27권 제2호, pp.193-200, 2000. 2.
- [7] 박형철, 전병환, "헤드 제스처 인터페이스를 위한 얼굴의 방향 및 눈의 깜빡임 정보 인식", 제14회 산학연 멀티미디어산업기술 학술대회 논문집, pp.320-324, 1999. 11. 19-20.