

# 입 모양 인식 기술의 비교 연구

## A study on lip-motion recognition algorithms

박한무 · 정진우

Han-Mu Park and Jin-Woo Jung

동국대학교 컴퓨터공학과

E-mail: lilees00@dongguk.edu, jwjung@dongguk.edu

### 요 약

얼굴 인식은 영상 처리 분야 중 대표적인 분야의 하나로, 지금까지 다양한 응용시스템이 개발됐다. 얼굴 인식은 눈, 코, 입 같은 얼굴의 특징들을 값으로 변환하고 각 특징 값들의 상관관계를 분석하는 방식으로 이루어지는데, 이 중에서 입은 형태 변화가 심하기 때문에 얼굴 인식에서는 특징 값으로 잘 이용되지 않는다. 반면, 표정 인식이나 화자 인식과 같은 특정 응용 시스템에서는 중요한 특징의 하나로 사용되고 있다.

입 모양을 인식한다는 것은 입술의 형태와 그 변화를 인식한다는 것을 의미하며, 이에 대한 연구가 많이 이루어지기는 했지만 음성 인식의 보조 수단으로 사용된 것이 대부분이다. 본 논문에서는 현재까지 제안된 입 움직임 인식 기술에 대해서 정리하고, 새로이 적용 가능한 응용 시스템에 대해 고찰해보고자 한다.

키워드 : 입 모양 인식, 실시간 처리, 영상 처리

### 1. 서 론

얼굴 인식은 영상 처리 분야 중 대표적인 분야의 하나로, 여러 가지 종류의 응용 시스템이 개발됐다. 최근에는 사용자의 표정을 인식하는 카메라가 상용화되었으며, 이외에도 다양한 응용 시스템이 개발 되고 있다.

얼굴 인식은 눈, 코, 입과 같은 요소들의 특징들을 값으로 변환하고 각 특징 값들의 상관관계를 분석하는 방식으로 이루어진다. 일반적으로 눈과 코가 특징 값으로 많이 사용되는데, 이것은 얼굴 영상에서 그 형태가 명확하며 위치 및 모양을 추출해내기 쉬운 편이기 때문이다. 반대로 입은 얼굴 영상에서 그 위치가 매우 명확함에도 불구하고 얼굴 인식 분야에서는 특징 값으로 잘 사용되지 않는데, 이는 다른 요소들에 비해 입은 형태 변화가 매우 심하기 때문이다. 반면에 얼굴 인식의 응용 분야인 표정 인식이나 화자 인식 같은 분야에서는 입이 매우 중요한 특징 값으로 사용된다. 대상이 웃고 있는지 울고 있는지, 말을 하고 있는지 아닌지 등을 판단하는데 있어서 입의 모양은 주요 판단 기준 중 하나이기 때문이다.

입 모양을 인식한다는 것은 입술의 형태와 그 변화 정도를 인식한다는 것을 의미한다. 변화 정도를 알기 위해서는 시간에 연속적인 여러 개의 영상이 필요하며, 인식에 사용되는 영상의 개수는 목적하는 응용 시스템과 알고리즘의 처리 속도에 따라 결정된다.

입 모양 인식을 위해 여러 가지 알고리즘들이 제안되었다. 본 논문에서는 최근 3년간 제안된 입 모양 인식 알고리즘들을 간단히 정리하고, 이러한 알고리즘들을 적용할 수 있는 응용시스템에 대해서 고찰해보고자 한다.

### 2. 입 모양 인식 알고리즘

입 모양을 인식하는 과정은 그림 1과 같이 크게 세 단

계로 구분할 수 있다. 전처리 과정은 얼굴영상을 입력으로 받아 입 영역을 추출하는 단계이고, 특징 값 추출 단계는 입 영역 영상에서 해당 영상만의 특징 값을 추출하는 단계이다. 이렇게 얻어진 특징 값들을 이용해 현재의 입 모양이 어떤 것인지를 판단하는 과정이 마지막 입 모양 인식 단계이다.

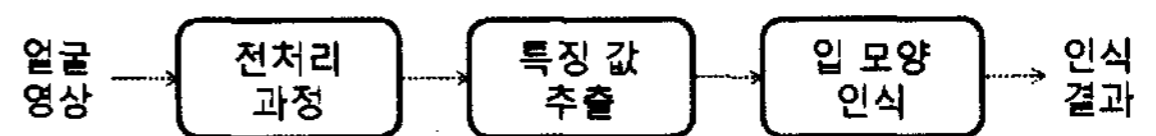


그림 1. 입 모양 인식 과정

Fig. 1. Lip-shape recognition process

#### 2.1 전처리 과정

전처리 과정은 입력 영상인 얼굴 영상에서 입이 존재하는 위치와 그 영역의 크기를 찾아내는 과정을 말한다. 전처리 과정 후 출력되는 입 영역을 대상으로 특징 값 추출 알고리즘들이 적용되며 입 모양의 특징을 검출하게 된다.

전처리 과정 후 생성되는 입 영역의 크기는 알고리즘에 따라 다르지만, 이 영역을 대상으로 특징 값 추출 알고리즘이 적용되기 때문에 영역 내부에는 실제 입의 모습이 모두 포함이 되어있어야 한다. 따라서 생성되는 입 영역의 크기는 최소한 실제 입의 크기보다 같거나 크게 설정되어야 한다.

입 영역을 추출하는 방법은 여러 가지가 있을 수 있는데, 예를 들어 [1]은 얼굴 영상의 RGB값을 이용해 이진화한 뒤, 템플릿 매칭으로 입 영역을 추출했으며, [2]에서는 연속된 영상들에서 변화가 발생한 모든 영역을 차영상을 이용해 구한 뒤, 각 영역을 대상으로 표 1과 같은 입술 움직임 특징 벡터 (해상도 320\*240 영상 기준)

와 유사성 정도를 비교해 가장 유사한 영역을 입 영역으로 추출했다.

표 1. 입술움직임 영상 특징요소별 데이터 [2]  
Table 1. Lip movement image features. [2]

Lip Feature Vector	가로 길이	세로 길이	가로 / 세로	면적율	평균 픽셀값	가로 위치	세로 위치
평균	5	20	3.95	0.37	16	1.03	0.68
표준 편차	2.18	6.88	0.87	0.07	3.92	0.18	0.1

### 2.2 특징 값 추출

특징 값을 추출하는 과정은 입 모양 인식 과정에서 매우 중요한 과정이다. 특징 값을 어떻게 정의하고, 어떤 방식으로 추출하느냐에 따라 성능이 크게 좌우되기 때문이다.

특징 값 추출 방법 역시 여러 가지 방법이 제안되었으며, 크게 두 가지로 구분해보면 입 영역 내부에 있는 픽셀이나 에지들의 분포를 특징 값으로 이용하는 방법과 입 영역 내부의 입술 윤곽선의 형태를 찾아내 그것을 특징 값으로 사용하는 방법이 있을 수 있다.

#### 2.2.1 영역 내 분포를 이용하는 방법

영역 내 분포를 이용하는 방법은 전처리 과정을 통해 지정된 입 영역내의 각 픽셀들을 색이나 엣지(Edge)들의 분포를 분석하고 그 정보를 이용해 입 움직임을 유추하는 방식이다.

Maycel-Isaac Faraj [4-5]는 입술의 움직임을 벡터들의 집합으로 표현하였다. 먼저 연속된 두 장의 영상을 비교해 입 영역 내의 점들이 얼마나 변화했는지를 계산한다. 그것을 벡터로 표현하면 그림 2와 같이 되는데, 이 움직임 벡터들을 그림 3처럼 6개의 영역으로 나누고, 각 영역에 속하는 벡터들을 화살표 방향 (0°, 45°, -45°)에 대한 스칼라 값으로 변경한다. 그렇게 되면 입 영역이 각 스칼라 값들의 2차원 행렬이 되는데, 이 행렬의 값을 다시 그림 4처럼 4등분해 클러스터링을 진행하게 된다. [4-5]에서는 fuzzy c-means 알고리즘을 이용해 클러스터링을 했으며, 클러스터링된 결과를 통계적으로 분석해 특징 값을 만들었다.

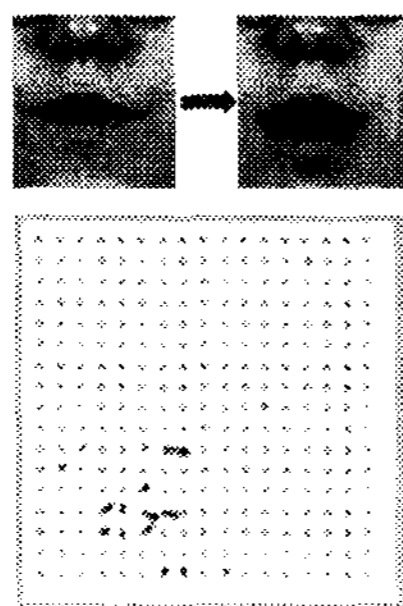


그림 2. 입술 움직임 벡터 [4]  
Fig. 2. The velocity vectors [4]

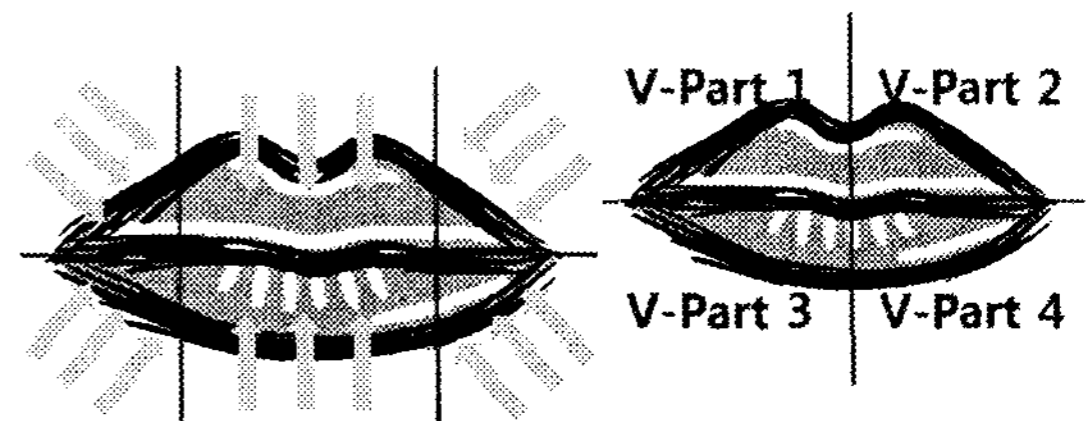


그림 3 움직임 벡터를 6개의 영역으로 분리 [4]  
Fig. 3. The Velocity vectors that are divided into six regions[4]

그림 4. 통계 분석을 위한 영역 분리 [4]  
Fig. 4. The estimated velocity vectors are divided into four regions[4]

[4-5]의 방법으로 입 움직임을 검출하는 것은 대상이 화자인지 아닌지를 검출하는 것을 목적으로 하고 있는데, 입 영상만을 가지고 화자 여부를 판단할 경우 78%의 정확도를 보였다.

#### 2.2.2 입술 윤곽선의 형태를 이용하는 방법

2.1절에서의 방법과는 다르게 입술의 윤곽선을 직접 찾고 이것을 특징 값으로 이용하는 방법도 제안 되었다.

[6-7]은 능동 형태 모델을 이용해 입술의 윤곽선을 추출했다. 능동 형태 모델이란 표준적인 모델을 설정하고 이 모델을 변경해가면서 물체의 실제 윤곽선을 찾는 방식이다. 만들어진 윤곽선은 PDM (Point Distribution Model)을 이용해 기준점들의 배열로 표현하였으며, [7]에서는 바깥쪽과 안쪽 입술의 형태를 총 41개의 점으로 표현하였다.

[6-7]의 방법은 검증을 위한 실험으로 수작업으로 표시한 입술의 영역과 알고리즘을 통해 만들어진 입술의 윤곽선을 비교하여 그 일치도로 알고리즘의 성능을 평가했고, 약 93%의 정확도를 보이는 것으로 나타났다.

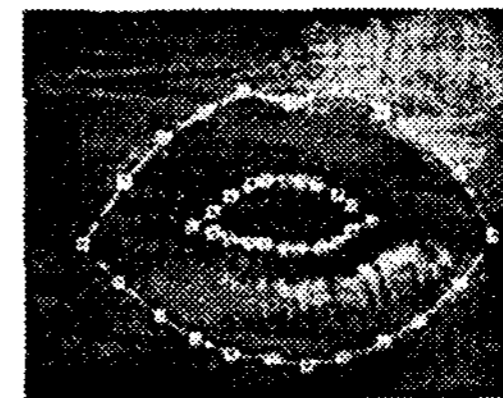


그림 5. 입술 형태[7]  
Fig. 5. Lip shape [7]

### 2.3 입 모양 인식

2.1절과 2.2절의 알고리즘들을 통해 특징 값들이 추출 되면, 미리 만들어진 분류기를 이용해 현재의 입 모양이나 움직임이 어떤 상태인지를 인식하는 것으로 인식 과정이 마무리 된다.

제안된 알고리즘에서는 베이지안 분류기 [8-9]나 SVM [4-5] 과 같이 잘 알려진 분류 방법을 통해서 입 모양을 인식하였다.

## 3. 적용 가능한 응용 시스템

지금까지 제안된 알고리즘들은 주로 음성 인식을 보조하기 위한 수단으로써 입 모양 인식 기술을 연구, 개발해왔다.[2,4-5,8-9] 음성 인식 시스템은 잡음이 많이 발생하며 이를 구분해내기 어려운데, 여기에 영상 처리를 접목함으로써 기존의 음성만을 이용한 시스템에 비해 인식

률이 크게 상승하는 효과를 얻어냈다.

입 모양 인식 기술을 적용 할 수 있는 분야는 이 외에도 여러 가지가 존재할 수 있다. 예를 들어, 3D 애니메이션 분야에 입 움직임 인식 기술을 적용한다면 3D 캐릭터의 입 모양을 좀 더 자연스럽게 표현 할 수 있을 것이다. 또한, 청각장애인의 조음 훈련이나 외국어 발음을 교정할 때, 입 모양 인식 기술을 이용해 현재 자신의 입 움직임과 고쳐야할 움직임을 화면에 표시해 발음 교정을 보조해주는 시스템도 생각해볼 수 있을 것이다.

IEEE Signal Processing Society, v.15 no.10, pp.2879-2891, 2006.

## 5. 결론

본 논문에서는 지금까지 제안된 알고리즘들을 대상으로 하여 각 방법을 비교해보고 새로이 적용 가능할 것으로 보이는 응용 시스템에 대해서 고찰해보았다.

그러나 각 알고리즘의 인식률은 각각 목적으로 하는 시스템이 달랐고, 실험 대상으로 했던 자료 역시 제각각이라서, 단순히 인식률만을 가지고 평가하는 것은 적합하지 않았다. 따라서, 추후 연구에서는 좀 더 자료를 조사하고, 모든 알고리즘들에 대해 객관적으로 평가할 수 있는 기준을 제시하는 것이 필요할 것으로 보인다.

## 참 고 문 헌

- [1] Lay Y.L., Tsai C.H., Yang H.J., Lin C.S., Lai C.Z., "The application of extension neuro-network on computer-assisted lip-reading recognition for hearing impaired," *Expert systems with applications*, v.34 no.2, pp.1465-1473, 2008.
- [2] 박준, 이영직, 김용규, 이수종, "입술움직임 영상신호를 고려한 음성존재 검출," *韓國音響學會誌* = The journal of the acoustical society of Korea, v.26 no.1, pp.25-31, 2007.
- [3] Shirasawa Yoichi, Nishida Makoto, Nishi Kenji, "A study on lip extraction due to fuzzy reasoning by using color information," *Electrical engineering in Japan*, v.157 no.1, pp.20-28, 2006.
- [4] Faraj M.I., Bigun J., "Audio-visual person authentication using lip-motion from orientation maps," *Pattern recognition letters*, v.28 no.11, pp.1368-1382, 2007.
- [5] Faraj M.I., Bigun J. "Synergy of Lip-Motion and Acoustic Features in Biometric Speech and Speaker Recognition," *IEEE Transactions on Computers*, v.56 no.9, pp.1169-1175, 2007
- [6] 장경식, 이임건, "형태계수의 Mixture Model을 이용한 입술 형태 표현과 입술 경계선 추출," *Journal of Korea Multimedia Society*, v.7 no.11, pp.1531-1539, 2004
- [7] 장경식, "에너지 최소화 기반 능동형태 모델을 이용한 입술 윤곽선 추출," *한국해양정보통신학회논문지* = The journal of the Korean institute of maritime & communication sciences, v.10 no.10, pp.1891-1896, 2006.
- [8] Cetingul H.E., Erzin E., Yemez Y., Tekalp A.M., "Multimodal speaker/speech recognition using lip motion, lip texture and audio," *Signal processing : the official publication of the European Association for Signal Processing (EURASIP)*, v.86 no.12, pp.3549-3558, 2006.
- [9] Cetingul H.E., Yemez Y., Erzin, E., Tekalp, A.M., "Discriminative Analysis of Lip Motion Features for Speaker Identification and Speech-Reading," *IEEE transactions on image processing : a publication of the*