

장애인의 입 모양 인식기법에 관한 연구

한성현, 권오상, *이용혁, 홍승홍
 인하대학교 전자공학과, *전양대학교 컴퓨터공학과

A Study on the Lip Shape Recognition Method for the Disabled

S. H. Han, O. S. Kwon, *E. H. Lee and S. H. Hong
 Department of Electronic Engineering, Inha University
 *Department of Computer Engineering, Konyang University

E-mail : g9711132@inhavision.inha.ac.kr

요 약

현재의 언어자동인식은 목소리의 음성 정보와 얼굴영상의 화상정보를 병행처리하여 초기의 음성 정보만을 처리한 경우보다 높은 인식률을 가져왔다. 이러한 연구는 언어의 인식에서 뿐만아니라, 신체가 부자연스러운 장애인에게는 간단한 의미 전달의 수단이나 각종 재활보조기기의 제어 등에도 매우 유용하게 활용될 수 있다.

음성정보처리를 배제한 영상정보처리 면에서, 본 논문의 대상이 되는 장애인은 신체의 부자연스러움으로 인해 처리되는 얼굴의 영상이 정상인과 같은 바른 자세가 되지 못하고, 기울어진다든지 틀어져서 기존의 방법들로는 정상인보다 현저히 낮은 입 모양의 인식률을 보인다.

본 논문에서는 기존의 방법들과는 달리 정색 표식을 이용한 색상분할법과 라벨링을 통해 입 모양을 인식하는 방법을 제안하였으며, 이는 장애인과 같은 기울어진 얼굴영상에 대해서도 정상인과 같은 입 모양 인식률을 나타내었다.

I. 서 론

사람의 언어를 자동으로 인식하는 기술은 1950년 대부터 연구가 활발히 진행되어 왔다. 초기에는 음성 자체의 처리에 의존하였으나, 90년대에 이르러 처리할 수 있는 하드웨어의 급속한 발전에 힘입어 마이크로 폰을 이용한 음성정보와 CCD카메라 등과 같은 영상입력장비를 통한 화상정보를 병행처리할 수 있는 방법으로 발전하였다.

이러한 방법은 특수한 실험환경이 아닌 일상적인 환경하에서는 여러 소리 잡음의 존재로 인해 구분이 불가능한 화자(話者)의 언어인식에 음성 정보뿐만 아니라 화자의 표정이라든지, 입술의 변화 등의 영상정보를 동시에 처리함으로써 언어 인식률을 높이고 있다.^{[4],[5]}

그 일례로, 1996년 스탠포드 대학에서 행한 연구에서는 특정 발음 /da/, /fa/, /la/와 /ma/의 4가지

발음에 대해서 음성처리만 사용한 경우, 영상처리만 사용한 경우, 음성과 영상처리를 병행한 경우의 3가지 경우에 실험한 결과, 후자의 경우에서 94%의 가장 높은 인식결과를 보였다.^{[7],[8]}

이러한 음성 정보와 영상정보를 이용한 자동언어 인식은 신체가 부자연스러운 장애인(障閼人)에 대해 언어인식의 면보다는 상대방에게 간단한 의미를 전달하거나, 각종 재활보조기기의 입력매체로서 매우 유용하게 활용될 수 있다.

그러나, 현재까지의 연구들은 대부분 정상인을 대상으로 한 것들이었다. 그 중에서도 입 모양을 영상 처리하는 분야에 초점을 맞춘다면, 입 모양이 정상인과 다른 양상(樣相)으로 변화하는 장애인의 경우에는 기존의 방법과는 다른 새로운 방법의 영상처리기법이 요구된다.

이에, 본 논문에서는 대상으로 하고 있는 장애인이 뇌성마비(腦性癱瘓) 환자와 같이 손, 발 등 신체가 자유롭지 못하여 얼굴의 영상이 기울어지거나 틀어져서 입력되었을 경우, 기존의 방법으로는 입 모양의 인식률이 현저히 떨어지는 단점을 보완하기 위한 새로운 영상처리기법에 관한 연구이다.

II. 영상처리기법

1. 전체 시스템의 개요

본 논문에서 영상처리를 하기 위한 시스템은 <그림 1>과 같이 구성하였다.

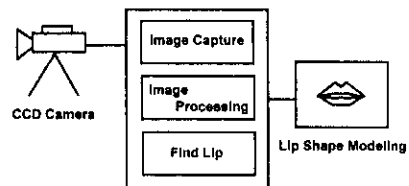


그림 1. 시스템의 구성

장애인의 얼굴은 설치된 CCD카메라와 이미지 그래픽을 통하여 초당 10프레임의 속도로 320*240 픽셀의 24비트 RGB영상으로 입력된다. 입력된 영상은 본 논문에서 제안한 영상처리기법을 통하여 입의 중심위치를 찾아낸다. 이렇게 찾아낸 입의 중심위치를 기준으로 입의 모양을 분석하여, 필요한 파라미터의 값을 측정한다. 이를 컴퓨터 상에서 모델링하였다. 여기서, 정상인과 같이 얼굴이 정면으로 똑바른 자세에서 입력된 영상이 아니어도 입 모양의 모델링이 가능한 영상처리기법을 제안하였다.

2. 얼굴 영상의 색상분할

사람의 얼굴을 RGB의 세 가지 색상별로 분리해보면 적색의 색상이 가장 많이 포함되어 있고, 다음으로 녹색, 청색의 순이다. 이러한 색상의 정보를 바탕으로 얼굴의 이마와 입 아래 부분에 청색의 9mm의 지름을 갖는 원형의 표식을 부착하게 되는데, 이는 본 논문에서 제안한 영상처리기법의 가장 중요한 정보를 제공하여 준다.

입력된 영상을 각각 적색, 녹색과 청색의 3가지 색상으로 분리하여 <그림 2>에 나타내었다.

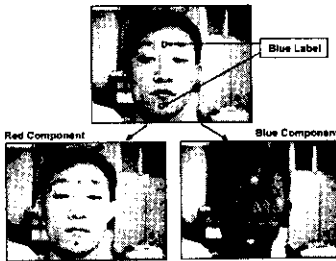


그림 2. 얼굴 영상의 색상분할

<그림 2>에서 RGB색상으로 분리된 좌측의 적색 요소는 전체 얼굴부위에서 높은 레벨(Gray Level)을 갖고, 청색의 표식에 대해서는 상대적으로 낮은 레벨 값을 갖는다. 우측의 청색 요소는 전체 얼굴에서는 낮은 레벨 값을 갖고, 표식에 대해서만 높은 레벨 값을 갖는다. 따라서, 이 두 영상의 레벨 값의 차를 이용하면 얼굴 부위를 검출해 낼 수 있다.

3. 표식의 검출과 얼굴의 중심선

얼굴에 붙인 표식의 검출로 얼굴의 기울임 정도에 대한 정보를 알 수 있으며, 입의 인식기법을 효과적으로 수행하기 위한 기준을 제공한다. 이를 위해서 <그림 3>의 방법을 사용하여 청색의 원형 표식을 검출하여 낸다.

먼저, 적색 요소의 영상과 청색 요소의 영상을 차를 구하고, 이를 임계치를 적용하여 얼굴 부위를 검출한 다음, 청색 요소의 영상에서 임계치를 두어 청색의 물체만 검출한 이진 영상에 얼굴 부위의 영역에서 청색 물체를 탐색하여 <그림 3>과 같은

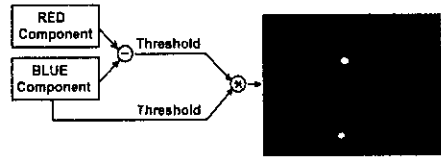


그림 3. 표식의 검출

검색 표식에 대한 이진 영상을 얻어낸다.

이 이진영상에 대해서 8-연결형 라벨링기법(Component Labeling using 8-Connectivity)을 적용하여 검출된 표식에 대한 각각의 중심좌표를 구한다. 이 두 중심좌표를 잇는 직선은 <그림 4>에서 보는 바와 같이 얼굴의 중심선이 된다.

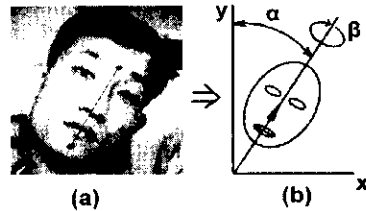


그림 4. 얼굴의 중심선

본 논문에서 대상으로 하는 장애인의 경우, 설치된 카메라를 정상인과 같이 똑바른 자세로 바라보지 못하므로, <그림 4>의 (b)와 같은 α , β 의 각도가 존재하게 된다.

여기서, α 는 y축을 중심으로 머리를 기울인 각도를 의미하며 두 표식의 좌표를 각각 (x_1, y_1) , (x_2, y_2) 라 하면, 다음과 같이 구할 수 있다.

$$\alpha = \tan^{-1}\left(\frac{|x_2 - x_1|}{|y_2 - y_1|}\right) \dots\dots\dots (\text{식 1})$$

각도 β 는 얼굴의 중심선을 중심으로 얼굴이 얼마나만큼 틀어졌는가를 나타내는 각도이다.

α 의 각도로 인하여 기울어진 얼굴의 영상은 y축을 중심으로 $-\alpha$ 만큼 회전시킴으로써 정상인과 같은 얼굴 영상으로 복원할 수 있지만, β 의 각도는 3차원 공간적인 변화가 2차원으로 투영되면서 발생하는 ROI(Region Of Interest)영상의 손실로 인하여 입 모양의 영상처리에 있어서의 인식률을 저하시킨다.

4. 히스토그램 분포를 이용한 입의 중심점 검출

본 논문에서는 입 모양을 알아내기 위해서 통계적인 히스토그램의 분포를 이용한다. <그림 4>과 같이 α 만큼의 각도로 기울어진 얼굴의 중심선을 따라 적색 요소의 영상에서의 히스토그램의 분포를 살펴보면 <그림 5>와 같다.

<그림 5>의 양끝에서는 청색의 표식이 낮은 레벨로 나타나며, 오른쪽 부근에 레벨의 변화가 크게 나타나는 입술의 중심(Center of Lip)이 검출된다. 이 입술의 중심좌표를 (x_c, y_c) 라고 하면, 중심선의

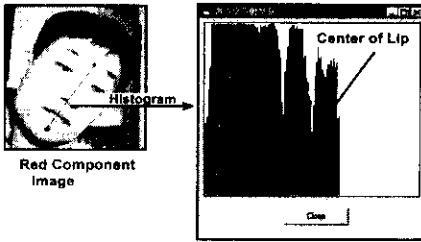


그림 5. 중심선의 히스토그램 분포

방정식은

$$y = m(x - x_c) + y_c \dots\dots (식 2)$$

이다. 여기서, m은 중심선의 기울기로 $\cot \alpha$ 이다.

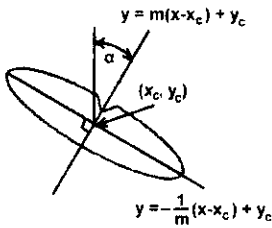


그림 6. 얼굴의 중심선과 법선

<그림 6>에서 보는 바와 같이, 입 모양의 정보를 얻어오기 위해서는 중심선과 수직인 법선이 필요하다. 이 법선의 방정식은

$$y = -\frac{1}{m}(x - x_c) + y_c \dots\dots (식 3)$$

이며, 이 법선을 기준으로 히스토그램을 분석하여 입의 모양을 인식한다.

5. 입 모양의 모델링을 위한 파라미터의 설정

입 모양을 모델링하기 위해서는 <그림 7>에서 표시한 a~e까지, 입술의 중심 등의 파라미터가 필요하다.

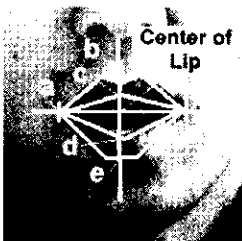


그림 7. 입 모양의 파라미터

a는 가로축상의 입의 거리를 나타내며, b와 c는 윗입술의 두께, d와 e는 아랫입술의 두께를 각각

나타낸다. 또한, c와 d로써 입의 벌림의 정도를 모델링할 수 있다.

이러한 파라미터에 의한 입 모양의 모델링은 입의 한쪽 면만을 이용하기 때문에 <그림 4>의 β 에 의한 얼굴의 뒤틀림에 대해서도 좋은 인식율을 보였다.

6. 파라미터의 측정

/e/로 발음한 경우, 중심선의 히스토그램을 이용하여 입술의 중심과 b~e까지의 파라미터를 측정된 결과를 <그림 8>에 나타내었다.

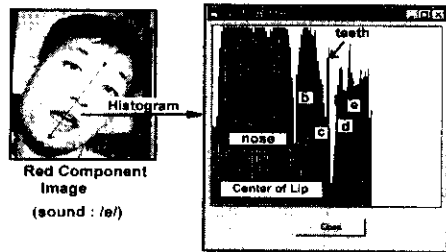


그림 8. 발음 /e/에서의 파라미터 측정

먼저, e파라미터를 측정하게 되는데, 아랫입술의 바로 아랫부분에서 입술의 레벨 값보다 상대적으로 작은 레벨 값이 감지되는 것을 이용한다. d파라미터는 레벨 값의 변화가 급격히 변화하기 바로 직전의 위치이다. 입의 중심으로부터 급격히 레벨이 높은 곳은 치아(齒牙)를 의미하는데, 치아의 존재여부는 발음의 종류, 파형자에 따라 달라질 수 있다. 치아를 제외하고, 입의 중심에서 급격히 레벨이 상승하는 위치에서 c파라미터를 결정하고, d와 c구간(아랫입술)에서의 레벨 값과 비슷한 구간을 벗어나는 위치를 b파라미터로 결정한다.

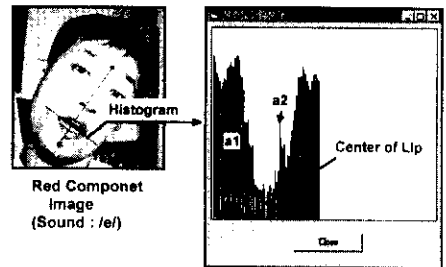


그림 9. 법선을 이용한 a 파라미터의 측정

마지막으로, a파라미터는 법선의 히스토그램의 분포를 분석하여 측정한다. a파라미터는 <그림 9>의 히스토그램을 분석하여 측정하며, 각각 a1과 a2의 두 개의 값을 얻어낼 수 있는데, <그림 4>의 β 의 각도가 존재하면 a1과 a2의 값이 다르다. 본 논문에서는 입술의 중심에서 가장 먼 거리의 값을 a파라미터의 값으로 취한다.

III. 입 모양의 모델링

/e/ 발음 시(그림 8) 실제 획득한 영상에서 측정된 파라미터의 좌표 값과 본 논문에서 제안한 알고리즘을 사용하여 측정된 파라미터의 좌표 값을 비교하여 <표 1>에 나타내었다.

<표 1>에서 보는 바와 같이 x축에서는 최대 5픽셀, y축에서는 최대 12픽셀의 오차를 보였다.

본 논문의 방법으로 측정된 파라미터를 이용하여 입 모양을 모델링하기 위하여 각각의 파라미터들간의 거리를 구하면, <표 2>와 같다.

표 1. 실제 영상과의 좌표 비교

파라미터	실제영상	제안알고리즘
a	(103, 167)	(108, 168)
b	(133, 171)	(128, 169)
c	(129, 175)	(125, 173)
Center of Lip	(124, 181)	(122, 177)
d	(119, 187)	(120, 180)
e	(115, 196)	(118, 184)

<표 2>의 거리를 이용하여 간단히 /e/에 해당하는 입 모양의 모델링을 직선으로만 표현하여 <그림 10>에 나타내었다.

표 2. 파라미터들간의 거리

구 간	거 리
a - Center of Lip	17
b - c	5
c - Center of Lip	5
Center of Lip - d	4
d - e	5

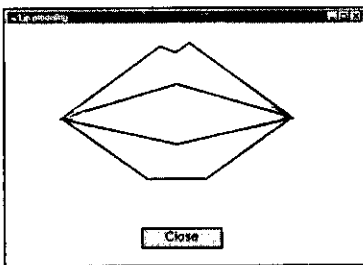


그림 10. 입 모양의 모델링

IV. 결 론

본 논문에서는 기존의 방법과는 달리 정상인이 아닌 장애인의 경우에도 입 모양의 인식에 효과적인 영상처리방법을 연구하였다. 이는 색상분할법을 기본 바탕으로 한 것으로, 이를 위해서 청색의 표식을 사용한다. 이 표식은 본 논문에서 가장 중요한 역할을 수행하지만, 영상처리를 하는데 있어서 반드시 부작을 해야 한다는 큰 단점으로 작용하고 있

다. 그러나, 이를 이용하여 기존의 방법들이 접근하기 힘들었던 장애인 얼굴 영상과 같은 기울림, 뒤틀림이 존재하는 얼굴 영상에 대해서도 효과적으로 수행할 수 있었다.

본 논문을 통해 얻어진 입 모양의 모델링은 입 모양의 변화양상이 정상인과 다른 장애인의 입 모양의 변화를 추적해 나갈으로써 장애인의 자동언어인식에 있어서 중요한 정보를 제공하여 줄 수 있다. 또한, 언어인식 분야에서뿐만 아니라, 장애인이 간단한 입 모양의 변화패턴(음성정보를 무시한다면 언어가 아닐 수도 있다.)만으로 재활보조기기를 제어할 수 있는 입력수단으로 활용될 수 있다. 이를 위해서는 입 모양의 인식뿐만 아니라, 간단한 명령어들을 인식할 수 있는 알고리즘이 요구된다.

참 고 문 헌

- [1] Gerhard X. Ritter and Joseph N. Wilson, "Handbook of Computer Vision Algorithms in Image Algebra", CRC Press, 1996
- [2] Randy Crane, "A simplified approach to Image Processing", Prentice Hall, 1997
- [3] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Addison Wesley, 1993
- [4] P. Jurlin, J. Luettin, D. Genoud and H. Wassner, "Acoustic-labial speaker verification", Proceedings of the First International Conference on Audio-and-Video-based Biometric Person Authentication, p.p. 319-326, 1997
- [5] J. Luettin, N. A. Thacker, and S. W. Beet, "Visual speech recognition using active shape models and Hidden Markov models", IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP'96), volume 2, p.p. 817-820, 1996
- [6] Li Nan, Shawn Dettmer and Mubarak Shah, "Lipreading Using Spatiotemporal Eigen Decomposition", Department of Computer Science Technical Report, 1995
- [7] Marcus E. Hennecke, K. Venkatesh Prasad and David G. Stork, "Automatic Speech Recognition System Using Acoustic and Visual Signals", 1996
- [8] K. Venkatesh Prasad, David G. Stork and Gregory J. Wolff, "Preprocessing video images for neural learning of lipreading", Ricoh California Research Center Technical Report, 1993