시각적 음성 인식을 위한 입술의 3차원 특징 분석

3D Lip Feature Analysis for Visual Speech Recognition

*고혜승[1,2], #윤인찬[1], 이연주[1], 황도식[2], 최귀원[1]
*H.S.Koh[1,2], #I.Youn(iyoun@kist.re.kr)[1], Y.J.Lee[1], D.Hwang[2], K.Choi[1]
[1]한국과학기술연구원 의공학연구소, [2]연세대학교 전기전자공학과

Key words : Stereo image processing, 3D data analysis, Visual speech recognition

1. 서론

의사소통은 사람의 의사나 감정의 소통으로, 인간이 사회생활을 하기 위해서 가지고 있어야 하는 필수적인 능력이다. 최근에는, 갈수록 복잡하고 다양해지는 기계로 인하여 사람간의 직접적인 소통이 아닌 기계와의 의사소통에 대한 관심이 증가하고 있다. 이러한 사람과 기계의 의사소통 방법 중 가장 대표적인 기술로는 구어를 매개로 하는 음성인식(Speech recognition)이 있다. 음성인식은 사람에게 익숙한 의사소통 수단인 음성으로 기계와 소통하는 기술로써, 별도의 학습이나 훈련이 없이도 기계를 손쉽게 사용할 수 있어 사용자에게 편리함을 제공할 수 있다[1]. 하지만 이러한 장점에도 불구하고 실제 음성인식 기술은 소음과 같은 배경잡음으로 인하여 인식률과 정확성이 저하된다[2]. 따라서 최근의 음성인식연구에서는 음성이 아닌 시각을 통한 음성인식연구가 많이 진행되고 있다. 시각을 통한 음성인식은 소음이 심각한 상황이나 음성 장애 환자들의 부정확한 음성에 대해서도 시각을 통하여 음성인식을 수행하기 때문에 배경잡음으로 인한 성능저하를 극복할 수 있다. 시각적 음성 인식(Visual Speech Recognition)은 입술의 위치를 파악하고, 입술의 움직임에 따라 특징점을 정확하게 추적하며, 피험자의 의도를 파악하기 위한 움직임 변화의 특징요소를 분석하는 것이 중요하다. 기존의 시각적 음성인식연구에서는 입술의 움직임을 분석하는 특징으로서 정면 영상의 특징인 입술의 너비, 높이와 측면영상의 특징인 입술의 깊이, 각도를 많이 사용하고 있다. 하지만 각 특징들의 조합이 음성인식의 향상정도에 어떠한 영향을 미치는지에 대하여 자세히 알려진 바가 없다. 따라서 본 논문에서는 스테레오 카메라를 이용한 3차원 입술 모양 추적 시스템[3]을 통해 인식의 정확도를 향상시킬 수 있는 입술 움직임 특징을 추출하고 분석한다.

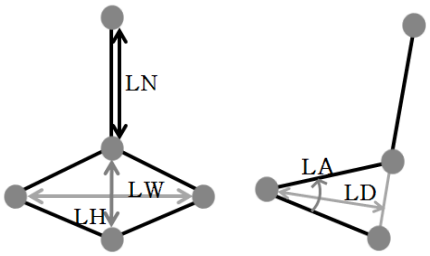
2. 3차원 입술 움직임 특징

본 논문에서는 스테레오 카메라를 이용하여 2차원적인 수평적 관계의 특징뿐만 아니라 깊이를 나타내는 제3의 축을 도입한다. 3차원 입술 좌표는 제3의 축을 지니고 있기 때문에, 얼굴의 회전에서도 스테레오 영상에서 동일한 특징점이 존재한다면 입술의 형태를 복원할 수 있으며 빠른 영상처리가 가능하다는 강점이 있다. 3차원 복원을 통해 획득한 입술 특징점 좌표는 시각적 음성 신호 데이터를 정확하게 분류하기 위한 분류 매개 변수를 획득하기 위하여, 그림 1과 같이 특징점의 기하학적 관계를 통해 5가지의 입술 움직임 특징을 추출한다. 5 가지의 입술 움직임 요소는 정면영상의 요소인 입술 너비(Lip Width, LW)와 높이(Lip Height, LH), 측면영상의 요소인 깊이(Lip Depth, LD)와 각도(Lip Angle, LA), 본 연구에서 제안하는 입술 위 끝점과 코 중심의 거리(Lip to Nose, LN)로써 이루어져 있다. 제안한 입술의 위 끝점과 코 중심의 거리는 입술의 돌출과 돌출 후 압축의 과정을 보이는 발음에 대하여 구분이 가능하게 하는 특징을 갖는다.

움직임 특징은 실험자의 앉는 위치에 따라 다르게 나타나며, 입술 움직임 특징의 크기 역시 변화한다. 따라서 데이터를 정확하게 분류할 수 있는 분류 매개 변수를 얻으려면, 발음의 시작과 끝 상태 사이의 변화를 계산하고, 초기 입술 위치에 정규화 해야 한다. 분류 매개 변수인 입술의 움직임 특징 요소를 정규화하기 위하여, 계측이 시작되는 시점인 침묵 상태와 완전히 발음을 완료한 상태를 구한 후, 식 (1)와 같이 계산한다. 정규화 된 입술 움직임 특징

요소는 계측을 시작하는 침묵상태(P_{mute})에 대한 발음과정($P_{utterance}$)의 특징변화율(P_r)을 나타내어, 사용자의 계측 위치나 포즈의 변화에도 강인하게 측정할 수 있다.

$$P_r = \frac{P_{utterance}}{P_{\mu te}} - 1 \quad (1)$$



(a) (b)

Fig. 1 3D Lip Feature: (a) Front image, (b) Sagittal image.

3. 결과

본 연구에서는 4명(남2, 여2)의 실험자에 대한 모음 5개(/a/, /e/, /i/, /o/, /u/)의 데이터 중 10개의 학습데이터와 5개의 테스트 데이터를 Support Vector Machine (SVM) 분류 알고리즘 통해 입술 움직임 특징별로 데이터를 분류하고, 표 1과 같이 입술 특징 비교를 수행하였다. 실험 결과, 정면영상 특징(LW, LH)의 경우 76.67%를, 측면영상특징(LD, LA)의 경우 65.83%를, 정면과 새로운 요소(LW, LH, LN)의 조합의 경우 68.33 %를, 측면과 새로운 요소(LD, LA, LN)의 조합의 경우 72.50%를, 정면과 측면의 조합(LW, LH, LD, LA)의 경우 85.83%를, 5가지의 모든 요소(LW, LH, LN, LD, LA)를 조합한 경우 87.50%의 평균 분류 정확성을 나타내었다. 따라서 5가지 요소 중 입술 움직임의 데이터 분류를 가장 잘 나타내는 요소는 5가지의 모든 특징을 포함하는 경우라는 것을 알 수 있다. 또한 정면 요소는 측면 요소보다 10.84%정도 정확성이 더 높지만, 정면과 측면 요소의 조합 결과와 비교하면 9.16% 낮다. 이러한 결과를 통해 일반적으로 사용되는 정면의 데이터보다 제 3축을 도입한 정면과 측면의 데이터 조합이 더 많은 정보를 가지고 입술의 움직임을 정확히 분류함을 알 수 있다. 뿐만 아니라 정면과 측면 요소와 새로운 요소를 조합한 결과를 비교하였을 때, 새로운 요소의 추가로 1.67%의 정확성 향상이 나타남을 알 수 있다.

Table 1 3D Lip Feature Analysis

(%)	LW, LH	LD, LA	LW, LH, LN	LD, LA, LN	LW, LH, LD, LA	LW, LH, LN, LD, LA
mute	100.00	100.00	100.00	100.00	100.00	100.00
a	75.00	35.00	60.00	50.00	60.00	65.00
e	85.00	35.00	65.00	45.00	80.00	85.00
i	75.00	50.00	50.00	75.00	80.00	85.00
o	75.00	80.00	70.00	70.00	95.00	95.00
u	55.00	95.00	65.00	95.00	100.00	100.00
평균	**76.67**	**65.83**	**68.33**	**72.50**	**85.83**	**87.50**

4. 결론

본 논문에서는 스테레오 카메라를 통한 3차원 입술 특징점 추출 시스템[3]을 통해, 인식의 정확도를 향상시킬 수 있는 5가지의 입술 움직임 특징을 추출하고 분석하는 방법을 제안하였다. 실험 결과, 5가지의 특징요소는 시각적 음성인식 데이터의 분류 정확성 향상에 도움을 주며, 5가지의 특징요소로 입술 움직임을 정확하게 분류하고 인식할 수 있다. 따라서 앞으로 시각적음성인식과정에서 본 논문의 5가지의 입술 요소를 통하여 향상된 정확성을 제공할 수 있을 것으로 예상된다.

후기

이 연구는 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단-공공복지안전사업의 지원을 받아 수행된 연구임 (No. 2012-0006551).

참고문헌

1. 최은정, “음성인식 기술의 재발견”, 삼성경제연구소 SERI 경영노트, **117**, 2011.
2. Gong, Y., “1. Speech recognition in noisy environments: A survey”, Speech Communication, **16**, 261-291, 1995.
3. 고혜승, 한성민, 추준욱, 박성희, 최재봉, 최귀원, 황도식, 윤인찬, "스테레오 카메라를 이용한 3차원 입술 모양 추적 시스템 개발," 한국 정밀공학회 춘계학술대회, 979-980, 2011.