Journal of Korea Multimedia Society Vol. 22, No. 8, August 2019(pp. 852-859)
https://doi.org/10.9717/kmms.2019.22.8.852

베이지안 분류 기반의 입 모양을 이용한 한글 모음 인식 시스템

김성우†, 차경애††, 박세현†††

Recognition of Korean Vowels using Bayesian Classification with Mouth Shape

Seong-Woo Kim†, Kyung-Ae Cha††, Se-Hyun Park†††

ABSTRACT

With the development of IT technology and smart devices, various applications utilizing image information are being developed. In order to provide an intuitive interface for pronunciation recognition, there is a growing need for research on pronunciation recognition using mouth feature values. In this paper, we propose a system to distinguish Korean vowel pronunciations by detecting feature points of lips region in images and applying Bayesian based learning model. The proposed system implements the recognition system based on Bayes' theorem, so that it is possible to improve the accuracy of speech recognition by accumulating input data regardless of whether it is speaker independent or dependent on small amount of learning data. Experimental results show that it is possible to effectively distinguish Korean vowels as a result of applying probability based Bayesian classification using only visual information such as mouth shape features.

Key words: Lip-reading, Bayesian Classifier, Korean Vowels, Mouth Shape

1. 서 론

IT기술의 발달로 인해 스마트 기기의 사용이 보편화되어지면서, 카메라를 이용한 영상 정보를 활용하는 다양한 어플리케이션이 개발되고 있으며, 직관적인 인터페이스의 제공을 위해서 영상에서 보이는 사람의 입 모양을 이용한 발음 인식에 대한 연구가 이루어지고 있다[1-6]. 발음의 인식을 위해서는 음성 정보의 분석이 필요하지만, 실제 환경에서 발생하는 잡음 요소로 인해서 제약을 받는 경우가 많기 때문에 보다 향상된 인식 결과를 얻기 위해서 영상 정보를 동시에 활용하는 연구뿐만 아니라, 시각적 특성만을 사용하여 발음을 인식하는 연구가 제안되고 있다[2-5]. 최근에는 얼굴 인식 학습 모델이 적용된 딥러닝 기술을 이용하여 입 모양 특성을 검출하여 발음 인식 등의 영역에 적용한 연구가 늘어나고 있다[2-4]. 하지만 딥러닝 알고리즘은 대용량의 학습 데이터를 확보해야 하며, 학습 모델 구축 과정의 시간적 요소와 개인화된 서비스에서 맞춤형 학습 모델을 적용 가능한 활용도의 측면에서 제한 사항이 있다.

※ Corresponding Author : Kyung-Ae Cha, Address: (712-714) Daegu Univ. Kyongsan Campus, Jillyang-eup, Gyeongsan-si, Gyeongsangbuk-do, Korea, TEL : +82-53-850-6641, FAX : +82-53-850-6629, E-mail : chaka@daegu.ac.kr
Receipt date : July 19, 2019, Approval date : Aug. 7, 2019

† School of Computer & Communication Engineering, Daegu University (E-mail : sungwoo6@daegu.ac.kr)
†† School of Computer & Communication Engineering, Daegu University
††† School of Computer & Communication Engineering, Daegu University (E-mail : sehyun@daegu.ac.kr)
※ This research was supported by the Daegu University Research Grant.

본 논문에서는 베이지안 분류에 의한 시각 정보만을 이용한 한글 모음 발음 인식 시스템을 개발하고자 한다. 제안하는 시스템은 얼굴의 특징점에서 입 모양 특징 벡터를 이용하여 한글 모음을 구분할 수 있는 실시간 학습 모델이다. 또한 특징 파라미터의 사전 확률을 계산하여 적은 수의 데이터만으로도 높은 확률의 결과를 도출할 수 있는 기법인 베이지안 이론[7-9]에 기반을 둔 알고리즘을 구현하여 적은 학습 데이터만으로도 화자 독립이거나 종속인 경우에 상관없이 입력 데이터의 축적을 통해서 발음 인식 확률을 향상시키는 시스템을 개발한다. 이를 통해서, 기존의 얼굴 특징점 인식 알고리즘으로 자주 사용되어진 SVM(Support Vector Machine)이나 CNN(Convolutional Neural Network) 기반 딥러닝 알고리즘에 비해 복잡한 계산을 요구하지 않고, GPU 등의 고성능 하드웨어에 사양에 구애받지 않을 수 있다.

본 논문의 2장에서는 영상 정보에서 입 모양의 특징을 이용한 발음 인식에 관한 기존의 연구들을 설명한다. 3장에서는 한글 모음 인식을 위해서 입 모양 특징점을 정의하고 이를 베이지안 분류기에 적용하여 발음을 인식하는 제안 방법에 대해서 기술한다. 4장에서 실험 결과를 보이고 5장에서 결론을 맺는다.

2. 관련 연구

영상에서 얼굴의 특징점을 인식하여 화자 구분이나 음성 인식 등에 관한 연구가 있으며, 특히 잡음 환경에서 음성 인식의 효율성을 높이기 위해서 음성과 영상 정보를 결합한 멀티 모달 시스템과 오디오-비주얼 음성 인식(AVSR, Audio-Visual Speech Recognition) 시스템이 제안되었다[1,10]. 이러한 시스템은 음성 정보의 분석이 필요하여 여전히 잡음 영향을 배제할 수 없어 음성 정보의 품질에 따라 인식률이 많이 달라질 수 있다.

이러한 문제점을 보완하기 위해서 영상 정보만을 이용하여 발음을 인식하기 위해서 입 모양 특징을 분석하여 발음 교정 등의 어플리케이션에 활용하는 시스템을 개발한 연구가 이루어졌다[2-5]. 이 연구들에서는 실험 단어를 발성한 입 영역의 특징 벡터 검출을 위한 과정으로 CNN 등의 알고리즘을 사용하거나[2,3], 입술 영역에 주성분 분해법(PCA, Principal Component Analysis)을 적용하여 특징을 추출하며[5], 입 모양 인식을 위해서는 HMM(Hidden Markov Model)이나 SVM 등을 사용하고 있다[3-5].

발음의 정확도를 시각 정보로 분석하는 시스템[4]에서는 dlip[11]의 얼굴 특징 랜드마크(Landmark)를 이용하여 입 모양의 특징점을 검출하고 표준이 되는 발음 모양과의 유사도를 검사하는 RNN(Recurrent Neural Network)기반 모델을 적용하였다.

한편, 시각 정보만으로 발음을 인식하기 위해서는 영상에서 얼굴과 입의 영역을 정확히 검출하는 것이 매우 중요한 요소이다[12]. 입 모양을 시각적으로 분석하여 발음 구간이나 해당 음성 정보로의 변환 등을 실험한 연구로 사람의 눈 위치에 기반하여 입의 위치를 찾고 주변 영역의 밝기 변화 등과 같은 픽셀 기반으로 움직임을 검출하는 옵티컬 플로우를 이용하는 기법들이 있다[13-15]. 여기에서는 영상 잡음에 영향을 받을 수 있어, 여러 전처리 과정이 수행되었다.

영상에서 관심 영역 간의 밝기 차이를 이용하여 특징을 검출하는 Haar Cascade 방식[16]은 사람 얼굴을 인식하고 눈, 코, 입 등의 객체를 검출하는 방식으로 자주 활용되고 있다[17].

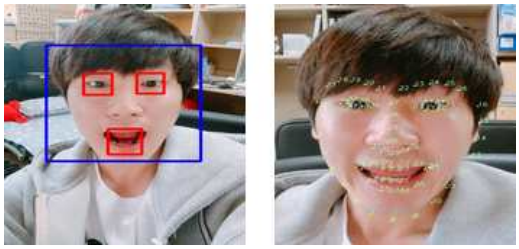
본 논문에서는 기존의 딥러닝 기반 얼굴 인식 모델을 사용하지 않고, 한글 모음 인식에 효과적인 특징 벡터를 설계하여, 기계학습의 확률 이론을 적용한 베이지안 학습 모델을 직접 구현하여 발음 인식에 효과적으로 활용될 수 있음을 실험 결과를 통해서 보인다.

3. 베이지안 분류를 이용한 한글 모음 발음의 인식

영상 정보를 이용한 발음의 인식을 위해서는 입 모양의 형태 변화를 특징 값들의 상관관계를 분석하는 방식으로 가능하다[18]. 이를 위해서 본 장에는 한글 모음 발음을 인식하기 위한 입 모양의 특징점을 정의하고 이를 검출하는 방식을 설명한다.

3.1 입 모양 특징점 검출

본 논문에서는 영상에서 얼굴을 감지하고 입 영역을 추출하기 위해서 Haar 알고리즘을 사용하였다. Fig. 1 (a)는 인식된 얼굴과 검출된 눈, 입을 표시한 것이다. Haar Cascades방식은 밝기 차를 이용한 간단한 검출기로 얼굴 등의 객체를 검출하는 속도를 향상시킬 수 있다. 검출된 영역에서 Fig. 1 (b)와 같이 dlib의 학습 모델을 이용하여 총 68개의 랜드마크를



(a) (b)

Fig. 1. Face Landmark Extraction: (a) Face Objects detected by Haar Algorithm, (b) Extracted Landmark.

추출한다. 특징점을 찾은 후 이미지 크기를 정규화하여 입 모양 특징점으로 정의한 요소들 사이의 거리를 계산한다. 따라서 영상에서의 얼굴 크기와 상관없이 일정한 특징점을 표현하는 속성 값을 구할 수 있다.

한글 발음에 있어서 입 모양 패턴은 대부분 모음에 의존하며 그 중에서도 기본이 되는 것은 단모음으로 입술의 움직임 정도에 따라서 발음이 구분된다[13,19,20]. 본 논문에서는 한글의 기본형 모음 중 'ㅏ[ɑ]', 'ㅣ[i]', 'ㅜ[u]', 'ㅗ[o]'의 네 개의 모음과 영어 모음에서도 사용되는 'ㅐ[æ](ㅔ[e])'의 총 다섯 가지의 모음을 구분할 수 있는 입 모양 특징점을 정의하고자 한다. 모음 발음이 시각 정보로 구분될 수 있는 것은 자음과 결합한 소리까지 인식하는데 있어서 매우 중요한 요소로 작용하기 때문이다.

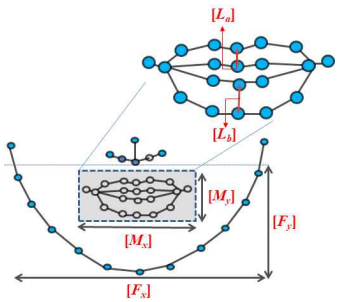
한글 모음 발음 시 입 모양은 위 입술과 아래 입술의 좌우, 위아래 움직임의 변화가 있으며, 둥근 모양의 정도가 달라지면서 발음이 구분된다[13,20]. 본 논문에서 정의한 입 모양 특징점은 Fig. 2와 같이 입 영역의 가로 길이[M_x], 입 영역의 세로 길이[M_y], 윗입술의 길이[L_a], 아랫입술의 길이[L_b]로 정의한다. 또한 인식률을 높이기 위해서, 입이 움직일 때 함께 변화를 보이는 턱과 볼을 사이의 거리를 이용하여 왼쪽 볼과 오른쪽 볼 사이의 거리[F_x], 입 영역의 상단과 턱의 하단 사이의 거리[F_y]를 측정하여 여섯 가지의 속성 값을 사용한다.

이와 같은 입 모양의 속성 값들은 발음마다 높은 확률로 나타나는 해당 특징점의 근사 값을 구할 수 있어, 모음 발음을 판별하는 척도가 될 수 있다. 예를 들어, 입을 옆으로 많이 벌리는 'ㅣ[i]'와 'ㅐ[æ](ㅔ[e])'의 경우에는 [M_x]와 [F_x]의 속성 값이 커지게 된다. 입을 아래으로 많이 벌리는 'ㅏ[ɑ]'와 'ㅜ[u]'의 경우는 [M_y]와 [F_y]의 속성 값이 커지게 된다. 또한 [F_x]의 값을 통해서, 'ㅏ[ɑ]'에서 'ㅣ[i]'나 'ㅐ[æ](ㅔ[e])' 발음으로 변할수록 양 볼이 넓어지는 특성을 반영하여 'ㅏ[ɑ]'와 'ㅐ[æ](ㅔ[e])'를 구별할 수 있다. 이는 베이지안 분류기의 학습 결과를 통해서 모음 발음을 구분할 수 있는 효과적인 특징점으로 사용됨을 볼 수 있다.

3.2 베이지안 분류 기반 한글 모음 인식

한글 모음의 발음 인식을 위해서 입 모양을 구분하기 위한 알고리즘으로 베이지안 분류 기법을 사용한다. 이를 통해서 입 모양 특징점으로 검출되는 속성 값들의 사전 확률(Prior Probability)과 사후 확률(Posterior Probability) 분포에 근거하여 훈련과 테스트 과정이 반복되면서 분류하고자 하는 모음의 확률 분포가 갱신되어가는 모델로 정의한다. 이는 적은 수의 학습 데이터로도 다섯 가지의 한글 모음 구분이 가능하고 화자 독립이나 화자 종속의 경우에 모두 적용할 수 있는 학습 모델이다.

베이즈 정리[8,9,21]는 식 (1)과 같으며, 여기서 $P(C|X)$는 사건 X가 일어난다고 가정했을 때 C가 일어날 조건부 확률로 이를 사후 확률이라고 한다. $P(C)$와 $P(X)$는 각 사건 C와 사건 X가 일어날 사전 확률이며, $P(X|C)$는 사건 C가 일어났을 때, 사건 X가



Attribute	M_x	M_y	L_a	L_b	F_x	F_y
Meaning	Width of the mouth area	Height of the mouth area	Thickness of the upper lips	Thickness of the below lips	Distance of the cheek edge	Distance between the top of mouth and the bottom of chin

Fig. 2. Definition and the Meaning of Feature Points.

발생할 확률을 의미하며 이를 우도(Likelihood)라고 한다.

$$P(C_i|X) = \frac{P(C_i)P(X|C_i)}{P(X)}, i = 0, 1, ..., n \quad (1)$$

본 논문에서는 식 (2)와 같이 한글 모음 다섯 개를 각각 사건 클래스 C_i로 정의하고, 입 모양 특징점으로 검출된 속성 값 $(M_x, M_y, L_a, L_b, F_x, F_y)$을 하나의 사건인 랜덤 변수 X로 정의한다.

$$P(C_i|X) = \frac{P(C_i)P(X|C_i)}{P(X)}, i = 0, 1, ..., 4$$
$$C_0 = [a], C_1 = [i], C_2 = [u], C_3 = [æ], C_4 = [o] \quad (2)$$
$$X = (M_x, M_y, L_a, L_b, F_x, F_y)$$

데이터 집합은 정규 분포를 따른다고 가정하면, 랜덤 변수 X가 속할 확률 분포인 한글 모음 중 하나를 결정하는 클래스 C_i의 확률 밀도 함수의 평균이 μ_i, 공분산이 $\sum_i$라고 할 때, 우도는 식 (3)이 된다.

$$P(X|C_i) = N(\mu_i, \sum_i) = \quad (3)$$
$$\frac{1}{(2\pi)^{d/2}\left|\sum_i\right|^{1/2}} exp(-\frac{1}{2}(x-\mu_i)^T \sum^{-1}(x-\mu_i))$$

랜덤 변수 X가 입력되면 C_i에 대해 사후 확률이 가장 큰 값으로 나타나는 클래스가 입 모양 특징점에 의해 분류되는 모음이다. 여기에서 식 (1)과 같이 사후 확률은 사전 확률과 우도의 곱으로 계산되므로, 최종적으로 우도가 최대인 클래스를 찾아 해당 모음 발음으로 분류한다.

4. 개발 및 실험 결과

4.1 실험 환경

제안하는 입 모양에 의한 한글 모음 인식 시스템의 검증을 위해 Table 1과 같은 환경에서 실험하였다. 사용된 영상 데이터는 20대, 30대, 그리고 50대 후반의 남, 여 10명의 발성 모습을 녹화한 총 500개로 구성된다. 500개의 데이터는 'ㅏ[ɑ]', 'ㅣ[i]', 'ㅜ[u]', 'ㅐ[æ](ㅔ[e])', 'ㅗ[o]' 의 다섯 개 모음 별로 각각 100개씩으로 이루어져 있다.

Table 1. Hardware Spec.

System	Specification
CPU	Intel Core i7-3770 3.40GHz
Graphic Card	Geforce GTX1060 3GB
RAM	16GB
OS	Windows 10 Pro
Tool	python 3.6.5
Library	OpenCV 2.4.10
Camera	IPhone XS Camera

학습 모델을 위한 훈련 이미지는 다소 일정한 조명 환경에서 촬영된 이미지를 사용하였으며 실험 데이터의 경우 밝은 실내와 어두운 실내에서 촬영한 이미지로 30초 이내의 발음 영상을 사용하였다. 단 어두운 실내의 경우 화자의 얼굴이 명확하게 보일 정도의 조명 상태를 유지하였다.

본 시스템은 딥러닝 알고리즘을 사용하지 않기 때문에 높은 수준의 하드웨어 사양을 요구하지 않는다. 따라서 하드웨어 구성은 Intel Core i7-3770 3.40GHz CPU와 Geforce GTX1060 3GB 그래픽 카드, 그리고 입력 영상의 해상도는 1090x1080로 구성하였다.

4.2 화자 종속 실험에 의한 입 모양 특징 벡터의 확률 분포

본 논문에서 제안한 시스템으로 먼저 동일인을 대상으로 하는 화자 종속인 경우의 한글 모음 인식률을 실험하였다. 같은 사람의 이미지 64장으로 훈련한 후, 110장의 이미지를 사용하여 테스트하였다. 훈련 이미지와 테스트 이미지는 다섯 가지의 발음이 같은 비율로 구성되어 있으며 훈련 데이터에서 무작위로 10회 반복하여 테스트 한 후 획득한 인식률의 평균을 계산하였다. 그 결과, 화자 종속의 경우 최대 94%의 발음 인식률을 보였다. 같은 화자의 경우 고유의 발음 모양을 가지기 때문에 발음에 따른 속성 값의 차이가 분명하고 한 가지 발음에 대한 속성 값의 변화가 크지 않아 인식률이 매우 높다.

실험에 의한 결과로, Fig. 3은 입 모양 특징점들에 대한 실제 획득된 속성 값의 확률 분포를 나타낸다. 여기서 입력 이미지는 500×700의 크기로 정규화 하였기 때문에 얼굴의 중앙 아래쪽에 위치하는 입 영역에서 추출되는 픽셀 값들은 대략 70 픽셀부터 최대 120까지의 값을 나타내고 있다. 각 속성의 픽셀 값은 얼굴 랜드마크에서 입 영역으로 정의된 특징들의 픽셀로 측정되는 거리를 의미한다.

Fig. 3을 통해 각 발음에서 해당 모음이 나올 수 있는 픽셀 값의 분포와 확률을 알 수 있다. 'ㅏ[ɑ]'의 경우 픽셀 값이 93 픽셀이 나올 확률이 높으며 'ㅜ[u]'

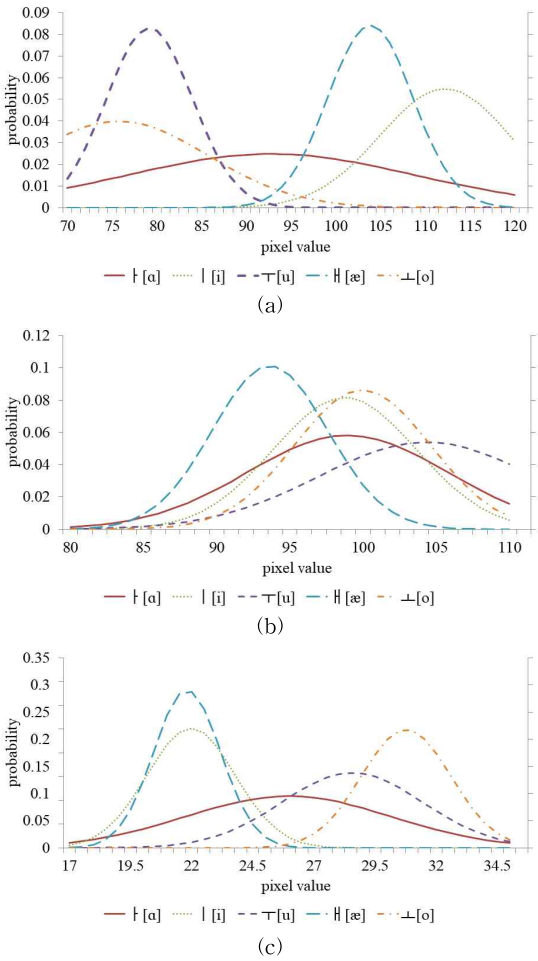


Fig. 3. Probability Distribution of Attribute Values for Same Person: (a) Attribute [M_x], (b) Attribute [M_y] and (c) Attribute [L_a].

의 경우 80 픽셀, 'ㅗ[o]' 발음의 경우 105 픽셀의 값이 나올 확률이 높다는 것을 나타낸다. Fig. 3 (a)는 입의 가로 길이인 [M_x] 값에 대한 훈련 이미지에 따른 확률 분포로써 각 발음별 분포도가 뚜렷이 구별된다. 특히 93 픽셀 값을 기준으로 'ㅜ[u]'와 'ㅗ[o]' 발음이, 'ㅐ[æ](ㅖ[e])'와 'ㅣ[i]' 발음이 확실히 구분되는 것을 볼 수 있다. Fig. 3 (b)의 [M_y]의 경우, 발음 별로 속성 값들이 근사 값이 많이 나타났지만 Fig. 3 (c)의 속성 [L_a]에서 'ㅐ[æ](ㅖ[e])' 발음과 'ㅗ[o]' 발음이 높은 확률로 분류되는 현상이 나타났다. 이는 베이지안 분류기의 입력인 랜덤 변수 X=($M_x, M_y, L_a, L_b, F_x, F_y$)의 각 속성 값들의 확률 값을 모두 곱하여 우도를 계산하기 때문에 발음을 구분하는데 더 높은 확률로 나타나는 속성에 의해서 발음 인식이 가능하다는 것을 보여준다.

4.3 한글 모음 발음 인식 실험 결과

위 4.1에서 언급한 총 500개의 이미지를 데이터로 실험한 결과, 전체적으로 약 85%의 발음 인식률을 나타내었으며 Table 2에서 보이는 바와 같이 'ㅏ[a]' 발음이 93%로 가장 높은 인식률을 나타내었다. 'ㅣ[i]' 발음이 85%로 두 번째로 높은 값을 보였으며, 가장 인식률이 낮은 것은 'ㅗ[o]' 발음으로 나타났다. 그 이유는 'ㅜ[u]'와 'ㅗ[o]'의 경우 속성 값의 확률 분포가 다소 비슷한 양상을 보이기 때문에 'ㅜ[u]'를 'ㅗ[o]'로 오인식되거나 'ㅗ[o]'를 'ㅜ[u]'로 오인식되는 경우가 있기 때문이다.

실험의 정확도를 향상시키고 dlib에서 얼굴 랜드마크를 100% 정확히 추출할 수 있도록 카메라를 정면으로 봤을 때 10도 내외의 일정 각도에서 얼굴을 인식하도록 하였다. 얼굴 영역은 머리카락의 경계가 되는 이마부터 턱까지 인식 영역 내에 진입하도록 하였으며, 이를 통해서 랜드마크 추출은 매 영상마다 정확한 값을 보인다고 할 수 있다.

Table 3은 본 논문과 유사하지만 ASM과 SVM을 사용한 연구[20]를 비교한 발음 별 인식률을 나타낸 표이다. Table 2와 Table 3의 결과 값은 랜드마크가 거의 100%의 정확도로 추출되는 환경에서 실험된

Table 2. Recognition result of each Korean vowel pronunciation

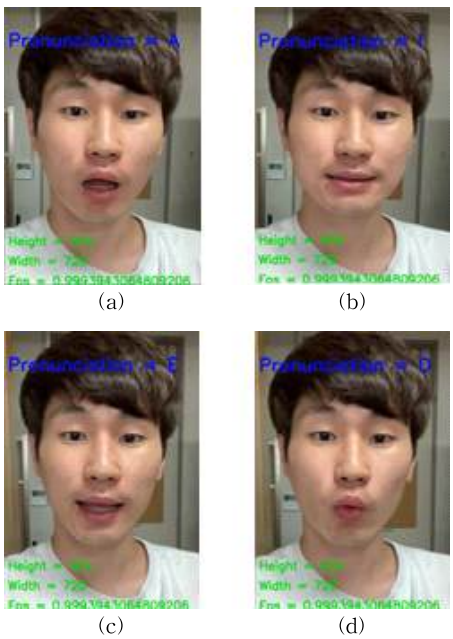
Vowel / Recogniton result	ㅏ[a]	ㅣ[i]	ㅜ[u]	ㅐ[æ] (ㅖ[e])	ㅗ[o]
ㅏ[a]	93%	0	0	0	0
ㅣ[i]	0	85%	0	16%	4%
ㅜ[u]	0	0	73%	0	17%
ㅐ[æ] (ㅖ[e])	0	15%	0	80%	0
ㅗ[o]	7%	0	27%	4%	79%

Table 3. Comparison of SVM[20] and proposed Bayesian classifier

Vowel	ㅏ[ɑ]	ㅣ[i]	ㅜ[u]	ㅐ[æ] (ㅔ[e])	ㅗ[o]
Bayesian	93%	85%	75%	80%	79%
SVM	67%	81%	84%	64%	73%

결과이며, SVM기반의 실험 결과 영상을 분석해 볼 때 이와 비슷한 인식 영역으로 실험되었다. 본 논문에서 제안한 방법에서는 'ㅏ[ɑ]' 발음이 93%로 가장 높은 인식률을 보였으며 SVM의 경우 'ㅜ[u]' 발음이 가장 높은 인식률을 보였다. 'ㅜ[u]' 발음을 제외한 모든 발음이 SVM을 사용한 경우보다 최대 26% 정도 높은 인식률을 보였으며, 이를 통해 기존 연구의 결과보다 향상된 인식 결과를 보인다는 것을 알 수 있다.

Fig. 4는 실시간 영상에서 프레임마다 각 발음을 추출한 결과이다. 영상은 아이폰 카메라로 촬영한 MPEG-4 동영상을 사용하였으며 인식된 발음을 영상의 상단부에 출력하였으며 영상에 대한 정보를 아래에 나타내었다. 영상은 30초의 'ㅏ[ɑ]', 'ㅣ[i]', 'ㅜ[u]', 'ㅐ[æ](ㅔ[e])', 'ㅗ[o]' 다섯 개의 모음 발음을 발화하는 영상이며, 입 모양에 맞는 발음을 정확히 화면에 나타나는 것을 볼 수 있다.



(a) (b) (c) (d)

Fig. 4. Snapshots of the Pronunciation detection results: (a) Pronunciation 'ㅏ[ɑ]', (b) Pronunciation 'ㅣ[i]', (c) Pronunciation 'ㅐ[æ]', (d) Pronunciation 'ㅗ[o]'.

5. 결 론

본 논문에서는 영상 정보에서 입 모양의 변화를 반영하는 특징값을 이용하여 한글 모음 발음을 인식할 수 있는 베이지안 분류 기반의 알고리즘을 구현하였다. 실험을 통해서 입 모양 특징 벡터의 확률 분포가 다섯 가지 한글 모음 발음을 구분할 수 있는 모수 분포로 갱신되어지며, 초기 학습 데이터의 양이 적더라도 모음 발음을 인식할 수 있음을 보였다.

또한 제안한 시스템은 입 모양의 특징점 검출을 위한 기존 연구에서 수행되어진 픽셀 기반의 이미지 처리 과정을 간소화하고, 딥러닝 기법에 비교하여 계산 복잡성이 낮아 학습이 시간이 오래 걸리지 않고 높은 사양의 하드웨어를 요구하지 않는다는 장점을 가진다.

이러한 시각 정보만을 활용한 발음 인식 연구는 여러 플랫폼에 적용이 가능하여, 소음이 심한 환경이나 청각적 불편함이 있는 사람들에게 효과적인 인터페이스를 제공할 수 있으며, 자동 자막 생성과 같은 연구에 도움이 될 것이다.

REFERENCE

[1] H.E. Çetingül, E. Erzin, Y. Yemez, and A.M. Tekalp, "Multimodal Speaker/Speech Recognition Using Lip Motion, Lip Texture and Audio," *Signal Processing*, Vol. 86, No. 12, pp. 3549-3558, 2006.

[2] Y. Xianoyi, *Lipreading Recognition of English Vowels Using Convolutional Neural Network and Recurrent Neural Network*, Master's Thesis of Chonbuk National University, 2017.

[3] Y.K. Kim, J.G. Lim, and M.H. Kim, "Lip Reading Method Using CNN for Utterance Period Detection," *Journal of Digital Convergence*,

Vol. 14, No. 8, pp. 233-243, 2016.

[4] D.Y. Lim, S.G. Kim, and K.T. Chong, "Development of a Real-time Lip Recognition for Improving English Pronunciation Using Deep Learning," *Journal of Institute of Control, Robotics and Systems*, Vol. 24, No. 4, pp. 327-333, 2018.

[5] C.G. Lee, E.S. Lee, S.T. Jung, and S.S. Lee, "Design and Implementation of a Real-Time Lipreading System Using PCA & HMM," *Journal of Korea Multimedia Society*, Vol. 7, No. 11, pp. 1597-1609, 2004.

[6] M.Y. Oh, Y.S. Jeong, and K.H, Park, "Driver Drowsiness Detection Algorithm Based on Facial Features," *Journal of Korea Multimedia Society*, Vol. 19, No. 11, pp. 1852-1861, 2016.

[7] S.W. Lee, J.W. Kim, J.H. Kim, and B.T. Zhang, "Neural Bayesian: The Paradigm of Prior Knowledge beyond Classical Bayesian Technique in Deep Learning," *Proceeding of Korea Computer Congress*, pp. 784-786, 2015.

[8] J.H. Choi, J.B. Kim, D.G. Kim, and K.W. Rim, "Bayesian Model for Probabilistic Unsupervised Learning," *Journal of Fuzzy Logic and Intelligent Systems*, Vol. 11, No. 9, pp. 849-854, 2001.

[9] G.D. Kleiter. "Propagating Imprecise Probabilities in Bayesian Networks," *Journal of Artificial Intelligence*, Vol. 88, No. 1, pp. 143-161, 1996.

[10] S. Lee, Y. Lee, H. Hong, B. Yun, M. Han, "Audio-Visual Integration based Multi-modal Speech Recognition System," *Proceedings of KIPS Fall Conference*, pp. 707-710, 2002.

[11] Dlib C++ Library, http://dlib.net/ (accessed March 10, 2019).

[12] K.T. Kim and J.Y. Choi, "Using Spatial Pyramid Based Local Descriptor for Face Recognition," *Journal of Korea Multimedia Society*, Vol. 20, No. 5, pp. 758-768, 2017.

[13] M.A. Lee, "A Lip-reading Algorithm Using Optical Flow and Properties of Articulatory Phonation," *Journal of Korea Multimedia Society*, Vol. 21, No. 7, pp. 745-754, 2018.

[14] S.M. Gyu, T.T. Pham, J.Y. Kim, and H.S. Taek, "A Study on Lip Detection Based on Eye Localization for Visual Speech Recognition in Mobile Environment," *International Journal of Fuzzy Logic and Intelligent Systems*, Vol. 19, No. 4, pp. 478-484, 2009.

[15] G.B. Kim, J.W. Ryu, and N.I. Cho, "Voice Activity Detection using Motion and Variation of Intensity in The Mouth Region," *Journal of Broadcast Engineering*, Vol. 17, No. 3, pp. 519-528, 2012.

[16] P. Viola and M.J. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 511-518, 2001.

[17] W. Hwang, "Research Trends in Deep Learning Based Face Detection, Landmark Detection and Face Recognition," *Broadcasting and Media Magazine*, Vol. 22, No. 4, pp. 41-49, 2017.

[18] H.M. Park and J.W. Jung, "A Study on Lip-motion Recognition Algorithms," *Proceedings of KISS Spring Conference*, Vol. 18, No. 1, pp. 268-270, 2008.

[19] Y.D. Lee, C.S. Choi, and K.S. Choi, "Lip Shape Synthesis of the Korean Syllable for Human Interface," *The Journal of Korean Institute of Communications and Information Sciences*, Vol. 19, No. 4, pp. 614-623, 1994.

[20] Y.K Kim, *Lib Reading Algorithm Using Bool Matrix and SVM*, Master's Thesis of Chungbuk University, 2014.

[21] I.S. Oh, *Pattern Recognition*, Kyobobook, Seoul, Korea, 2008.

김 성 우

2018년 2월 대구대학교 정보통신공학부 학사
2018년 3월~현재 대구대학교 정보통신공학과 석사과정
관심분야 : 영상처리, 스마트어플리케이션, 인공지능, 딥러닝

차 경 애

1996년 2월 경북대학교 컴퓨터과학과 학사
1999년 2월 경북대학교 컴퓨터과학과 석사
2003년 8월 경북대학교 컴퓨터과학과 박사

2005년 3월~현재 대구대학교 정보통신공학부 교수
관심분야 : 멀티미디어처리, 스마트어플리케이션, 인공지능, 딥러닝

박 세 현

1995년 2월 경북학교 컴퓨터공학과 학사
1997년 2월 경북학교 컴퓨터공학과 공학석사
2000년 2월 경북학교 컴퓨터공학과 공학박사

2004년 3월~현재 대구학교 정보통신공학부 교수
관심분야 : 컴퓨터비전, 인공지능, 딥러닝