한국음향학회지 제34권 제4호 pp. 321~327 (2015)
The Journal of the Acoustical Society of Korea Vol.34, No.4 (2015)
http://dx.doi.org/10.7776/ASK.2015.34.4.321

pISSN : 1225-4428
eISSN : 2287-3775

음성인식기 성능 향상을 위한 영상기반 음성구간 검출 및 적응적 문턱값 추정

Visual Voice Activity Detection and Adaptive Threshold Estimation for Speech Recognition

송태엽, 이경선,* 김성수,** 이재원,** 고한석*†

(Taeyup Song, Kyungsun Lee,* Sung Soo Kim,** Jae-Won Lee,** and Hanseok Ko*†)

고려대학교 바이오마이크로시스템기술 협동과정, *고려대학교 전기전자전파공학부, **삼성전자

(Received April 16, 2015; accepted May 28, 2015)

초 록: 본 연구에서는 음성인식기 성능향상을 위한 영상기반 음성구간 검출방법을 제안한다. 기존의 광류기반 방법은 조도변화에 대응하지 못하고 연산량이 많아서 이동형 플렛홈에 적용되는 스마트 기기에 적용하는데 어려움이 있고, 카오스 이론 기반 방법은 조도변화에 강인하지만 차량 움직임 및 입술 검출의 부정확성으로 인해 발생하는 오검출이 발생하는 문제점이 있다. 본 연구에서는 기존 영상기반 음성구간 검출 알고리즘의 문제점을 해결하기 위해 지역 분산 히스토그램(Local Variance Histogram, LVH)과 적응적 문턱값 추정 방법을 이용한 음성구간 검출 알고리즘을 제안한다. 제안된 방법은 조도 변화에 따른 픽셀 변화에 강인하고 연산속도가 빠르며 적응적 문턱값을 사용하여 조도변화 및 움직임이 큰 차량 운전자의 발화를 강인하게 검출할 수 있다. 이동중인 차량에서 촬영한 운전자의 동영상을 이용하여 성능을 측정한 결과 제안한 방법이 기존의 방법에 비하여 성능이 우수함을 확인하였다.
핵심용어: 음성구간 검출, 지역 분산 히스토그램

ABSTRACT: In this paper, we propose an algorithm for achieving robust Visual Voice Activity Detection (VVAD) for enhanced speech recognition. In conventional VVAD algorithms, the motion of lip region is found by applying an optical flow or Chaos inspired measures for detecting visual speech frames. The optical flow-based VVAD is difficult to be adopted to driving scenarios due to its computational complexity. While invariant to illumination changes, Chaos theory based VVAD method is sensitive to motion translations caused by driver's head movements. The proposed Local Variance Histogram (LVH) is robust to the pixel intensity changes from both illumination change and translation change. Hence, for improved performance in environmental changes, we adopt the novel threshold estimation using total variance change. In the experimental results, the proposed VVAD algorithm achieves robustness in various driving situations.
Keywords: Voice activity detection, End-point detection, Local variance histogram
PACS numbers: 43.72.Ne, 43.72.Ar

I. 서 론

음성구간 검출은 입력신호로부터 음성구간과 비음성구간을 구별하는 기술로, 음성인식, 잡음 제거 등 다양한 응용분야에서 사용되고 있다. 특히, 음성인식 분야에서 고성능의 음성구간 검출 알고리즘[1]을 사용할 경우 비음성구간에 존재하는 다양한 잡음신호를 무시할 수 있고, 최적의 음성신호를 인식기에 전달할 수 있기 때문에 성능에 큰 영향을 미치게 된다. 하지만 음향 정보 기반 음성구간 검출 알고리즘(Audio Voice Activity Detection, AVAD)의 경우 충격음 등의 임펄스 노이즈에 매우 취약하기 때문에 차량 주행 중 상황과 같이 잡음이 심한 환경에서는 음향 정보만을 이용한 검출에는 한계가 있다. 이런

†**Corresponding author:** Hanseok Ko (hsko@korea.ac.kr)
Engineering building room 419, Korea University, 145 Anam-Ro, Seongbuk-Gu, Seoul 136-713, Republic of Korea
(Tel: 82-2-3290-3239, Fax: 82-2-3291-2450)

문제점을 해결하기 위한 방법으로 영상을 이용한 음성구간 검출 알고리즘이 활발하게 연구되고 있다. 영상기반 음성구간 검출 알고리즘은 촬영된 사용자의 얼굴 영상에서 입술의 움직임을 이용하여 음성구간을 검출하는 알고리즘으로, 화자의 얼굴 및 입술 영역을 찾는 검출부, 화자의 입술의 움직임을 정량화 하는 특징 추출부 그리고 특징값을 기준으로 음성구간 여부를 판단하는 검출부로 구성되어 있으며, 기존의 연구에서 실내에서 발화하는 화자에 대해 다양한 특징 및 판단 방법을 적용한 알고리즘이 제안되어 왔다. 초기의 연구는 분리된 입술 영역을 발화 시의 입술 영상과 비교하여 음성구간 여부를 판단하는 템플릿 매칭 방법이 사용되었다.[2] 하지만 조도변화에 민감하기 때문에 조명 및 환경에 따라 성능변화가 심한 문제가 있다. 광류(optical flow)를 이용한 방법[3,4,5,6,7]은 입술이 변화할 때의 픽셀값의 변화를 이용하여 모션백터를 산출한 후 정량화 하여 특징을 추출하는 방법으로 이 방법 역시 픽셀의 화소값의 변화를 이용한 방법이기 때문에 급격한 조도변화에 대응하지 못하는 단점이 존재하며, 연산양이 많기 때문에 차량환경 등 이동형 플렛홈에 사용되는 스마트 기기에 적용하는데 어려움이 있다. 카오스 이론 기반 방법[8,9]은 시간상에서 인접한 두 영상에서 입술의 움직임과 같이 비선형적으로 모델링되는 변화가 존재하는 경우 결합 분포에서 카오스 패턴이 나타나는 것을 이용하는 방법으로, 패턴의 복잡도를 정량화하여 입술의 움직임을 검출하는 방법이다. 이 방법은 전역적인 조도변화에 강인하게 입술 움직임을 검출할 수 있지만 차량 움직임에 의한 화자의 움직임과 입술검출기의 부정확성 등으로 음성구간 검출 과정에서 오검출하는 발생하는 문제점이 존재한다. 조도 변화에 강인한 특징 중 하나인 지역적 패턴 기반 방법은 각 화소에 인접한 화소집합에 대해 통계량을 추출하여 특징으로 사용하는 방법으로, 최근 제안된 방법 중 하나인 지역적 분산 히스토그램[10]을 이용한 음성구간 검출 방법은 각 화소집합에 대해 지역적 분산을 계산한 후 히스토그램을 구성하여 입술의 움직임에 따른 히스토그램의 변화를 특징으로 사용한다. 이 방법은 지역적 통계량을 추출하기 때문에 조도변화에 강인하고, 연산량이 적어 모바일 플렛폼에 적합한 방법이지만, 2가지의 문턱값을 사용하기 때문에 초기값 설정에 의해 성능이 변화하는 문제점이 있다. 본 연구에서는 기존 영상기반 음성구간 검출 알고리즘의 문제점을 해결하기 위해 지역분산 히스토그램[10,11]과 적응적 문턱값 추정 방법을 이용한 음성구간 검출 알고리즘을 제안한다. 제안된 방법은 지역 분산 히스토그램 특징의 조도 변화에 강인하고 연산속도가 빠른 특징을 유지하면서, 적응적 문턱값을 사용하여 급격한 조도변화 및 움직임이 큰 운전자의 발화를 강인하게 검출할 수 있음을 확인하였다. 제안한 알고리즘의 성능을 확인하기 위하여 이동 중인 차량에서 촬영한 화자의 동영상을 이용하여 기존 방법과 성능을 비교한 결과, 기존 방법에 비하여 제안한 방법이 성능이 우수함을 확인하였다.

II. 얼굴 및 입술영역 검출 방법

2.1 얼굴 및 입술영역 검출 방법

고성능의 영상기반 음성구간 검출을 수행하기 위해 본 연구에서는 화소의 구조적 패턴 및 Adaboost 훈련 알고리즘을 이용한 얼굴/입술 검출을 수행했다. 입력 영상의 특정 화소 위치에서 그 주변 화소 각각의 밝기 값(intensity)과 주변 화소들의 평균 밝기 값과의 비교를 통하여 특정 위치 화소에 대한 구조적 패턴을 표현할 수 있다. 본 논문에서는 Modified Census Transform(MCT) 특징을 적용하여 입력영상의 구조적 패턴을 정량화 하였다. MCT 특징은 3 × 3의 커널을 이용하여 특정 화소와 상하좌우 및 대각의 주변 8개 화소에 대한 구조적 패턴을 표현한다.[12] 화소 x 에서의 구조적 패턴은 Eq.(1)과 같이 정의할 수 있다.

$$I'(\boldsymbol{x}) = \otimes_{\boldsymbol{y}\in N'(\boldsymbol{x})}\ \zeta(\bar{I}(\boldsymbol{x}), I(\boldsymbol{y})), \tag{1}$$

N'(x)는 커널내의 화소들의 집합을 나타낸 것이며, $I(x)$는 화소 x 에서의 밝기 값, $\bar{I}(x)$는 $N'(x)$의 평균 밝기 값이다. ζ는 비교 함수로서, $\bar{I}(x) < I(x)$인 경우 1, 그렇지 않은 경우 0 값을 가진다. $\otimes$는 연결 연산자로서 커널 내 화소의 비교 연산 결과를 2진수 배열로 나열하고 10진수로 변환한다. 본 연구에서 사용한 정

사각형 커널은 9개 화소의 집합이고, $111111111_{(2)}$의 패턴은 가질 수 없으므로 특징 값은 총 2^9-1=511개의 종류를 한다.[12] 먼저 화소의 구조적 패턴 특징값 분류에 적합한 이진 약분류기(binary weak classifier)를 설계하고, Adaboost 훈련을 통해 약분류기의 조합을 통해 강분류기(strong classifier)를 생성한다. 훈련 과정을 통해 선택된 약분류기는 각각 신뢰도가 부여되며, 그 값을 누적시켜 조합된 강분류기의 신뢰도 색인표(Look-Up Table)를 만든다. 최종 강분류기는 Eq.(2)과 같이 정의된다.

$$H_i(\Gamma) = \sum_{\boldsymbol{x} \subset \Psi} h_{\boldsymbol{x}}(\Gamma(\boldsymbol{x})), \tag{2}$$

여기서 x는 각 화소의 주변 화소값이며, Γ는 Eq.(1)과 같은 구조적 패턴 특징값, Ψ 는 십자형 패턴의 화소 집합, h_x는 약분류기를 나타낸다. 얼굴 및 입술 검출 유무는 Eq.(3)과 같이 강분류기와 사용자가 설정한 문턱값(threshold)과의 비교를 통해 결정한다.

$$H_i(\Gamma) \le T_i, \tag{3}$$

여기서 T_i 는 i번째 단계의 강한 분류기의 문턱값이며, 검출률이 최대가 되도록 조절한다.

2.2 지역 분산 히스토그램 특징추출

입력영상에서 검출된 입술 영역에 대해 지역 분산 히스토그램을 계산하는 과정은 다음과 같이 나타낼 수 있다. Fig. 1과 같이 입술영역의 그레이 스케일 영상에서 중간 픽셀 값을 기준으로 주변 픽셀과 평균값의 차이를 통한 분산값을 구한다.

$$VAR_{P,R} = \frac{1}{P}\sum_{p=0}^{P-1}(g_p - \mu)^2, \text{ where } \mu = \sum_{p=0}^{P-1} g_p, \tag{4}$$

Eq.(4)에서 g_p는 중간 픽셀 위치에서 R 만큼 등거리에 떨어진 P개의 그레이 값들을 가리킨다. 각 픽셀에서 계산한 지역 분산의 히스토그램 H_i을 구하면 Fig. 2와 같이 입술의 움직임이 적은 영상 프레임은 높은 분산을 가지는 픽셀의 수가 적음을 알 수 있고, 화자가 발화하여 입술의 움직임이 발생하는 영상 프레임에

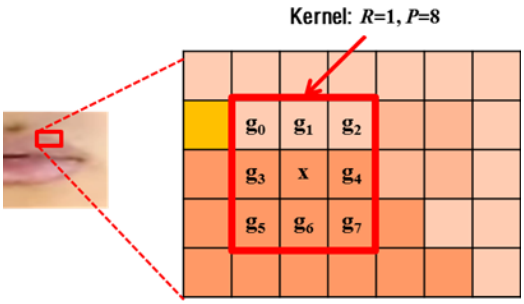

Fig. 1. Example of Local Variance Histogram.

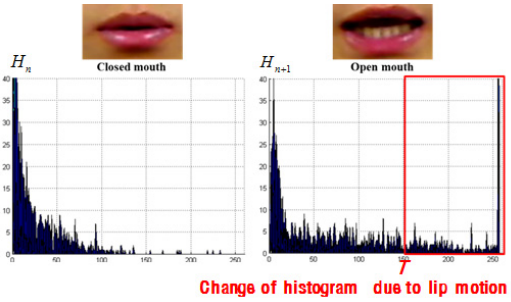

Fig. 2. Example of LVH chage from lip motion.

서 구한 히스토그램에 높은 분산을 가지는 픽셀이 많이 측정됨을 알 수 있다.

즉 문턱값 T_V이상의 분산을 가지는 픽셀수의 시간의 따른 변화량을 이용하여 음성구간 영상 프레임을 검출할 수 있다. 이를 이용한 지역 분산 히스토그램 특징 $v[n]$은 Eq.(5)와 같이 나타낼 수 있다.

$$v[n] = |c[n] - c[n-1]|, \; where\, c[n] = \sum_{i=T_r}^{M} H_n(i). \tag{5}$$

지역 분산 히스토그램 특징을 추출한 후 문턱값 T_f를 이용하여 해당 영상 프레임의 음성구간 여부를 판단하게 된다. 이는 Eq.(6)과 같이 나타낼 수 있다.

$$F[n] = \begin{cases} 1, & v[n] > T_f \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Eq.(6)에서 $F[n]$은 음성구간 여부를 나타내는 변수로, 1값을 가질 때 음성구간, 0 값을 가질 때 비음성구간을 나타낸다. 매 시간 입력되는 영상에서 음성구간을 판단하는 경우 발화시 입술의 움직임에 따라 일부 음성 구간에서 순간적으로 비음성 구간으로 판

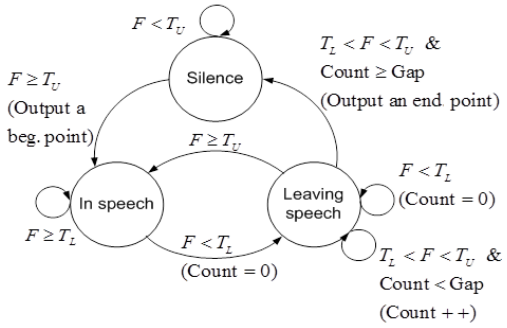

Fig. 3. State transition diagram for VVAD.

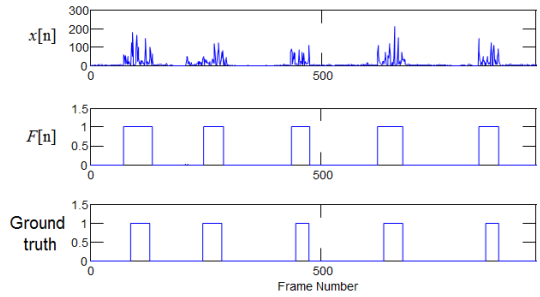

Fig. 4. Result of state transition model for VVAD.

단할 수 있으므로, $F[n]$에 대해 상태 천이의 동작[13]을 통하여 최종 결과를 도출한다. Fig. 3는 상태 천이 모델이다. Fig. 3에서 Silence는 비음성 구간을 나타내고 In speech는 음성 구간을 나타낸다. Leaving speech는 음성 구간이지만 비음성 구간으로 변할 수 있는 단계이다. T_L은 낮은 문턱치(lower threshold), T_U는 높은 문턱치(upper threshold), Gap은 끝나는 점을 결정하기 위한 허용치로써 실험적으로 정하는 상수이다. 단, T_U는 항상 T_L보다 커야 한다. 위 상태 천이 모델을 이용하면, $F[n]$이 T_U보다 작으면 음성이 없는 비음성 구간(Silence)으로 판단 한다. $F[n]$이 T_U보다 커지면 음성이 시작된 것으로 보고 그 부분을 시작점(In speech)으로 잡는다. $F[n]$이 T_L보다 작아지면 아직 음성 구간이긴 하지만 비음성 구간으로 바뀔 가능성(Leaving speech)이 있다고 간주하고, *Count* ($F[n]$이 T_L와 T_U 사이에 있는 경우 연속적으로 그 사이에 있는 횟수)를 0으로 잡는다. *Count*가 *Gap* 보다 작으면 Leaving speech로 판단하고, *Count*가 *Gap* 보다 크면 Silence로 판단한다. Silence로 판단되는 그 프레임이 끝나는 점이 된다. $F[n]$이 T_L 보다 작아지면 *Count*를 *0*으로 잡고 Leaving speech 단계를 유지한다. 그리고 $F[n]$이 T_U 보다 커지면 다시 In speech 구간으로 돌아간다. 음성의 시작점과 끝나는 점을 검출하기 위해 프레임 에너지에 위 기법을 적용한 결과를 Fig. 4에 나타내었다. Fig. 4는 주행중인 운전자를 촬영한 영상에서 입술을 검출한 후 입술의 특징값에 상태 천이 모델을 적용하여 끝점 검출한 결과이다.

2.3 적응적 문턱값을 이용한 음성구간 검출

본 논문에서는 지역 분산 히스토그램 특징을 이용하여 음성/비음성 영상프레임을 검출하기 위해 적응적 문턱값 추정 방법을 사용하였다. Otsu[14]의 연구를 응용하여 음성/비음성으로 구분되는 특징에 대해 클래스간의 분산을 최대화 시키는 문턱값을 추정한다. 순차적으로 입력되는 영상 중 n번째 프레임에서 얻은 특징을 $v[n]$ 이라 할 때, 이전 프레임으로부터 얻은 L 개의 특징에서 값이 i 인 특징값의 수를 f_i, 특징의 총 수를 N이라고 할 때, 특징이 i 값을 갖을 확률은 $p_i=f_i/N$ 이 된다. 특징이 음성구간과 비음성구간 두 개로 나누어 질 수 있다면, 특징들을 폐구간 $[1,...,T]$에 속하는 클래스 C_1과 폐구간 $[T, ... L]$ 에 속하는 C_2 클래스로 나눌 수 있다. 이때 클래스 C_1과 C_2의 평균은 Eqs.(7)과 (8)로 나타낼 수 있다.

$$\mu_1(t) = \sum_{i=1}^{T_f} ip_i/\omega_1(t), \tag{7}$$

$$\mu_2(t) = \sum_{i=T_f+1}^{L} ip_i/\omega_2(t)3. \tag{8}$$

전체 특징의 평균은 Eq.(9)와 같이 나타낼 수 있다.

$$\omega_1\mu_1 + \omega_2\mu_2 = \mu_T$$
$$\text{where } \omega_1(T_{T_f}) = \sum_{i=1}^{T_f} p_i,\ \omega_2(T_{T_f}) = \sum_{i=T_f+1}^{N} p_i. \tag{9}$$

전체 특징의 분산은 Eq.(10)과 같이 나타낼 수 있다. 이때 클래스 C_1과 C_2에 대하여 클래스 내의 분산(within-class variance)을 σ_w^2, 클래스 사이의 분산(between-class variance)을 σ_B^2라 할 때 전체 특징값에 대한 분산은 Eq.(11)과 같이 나타낼 수 있다.

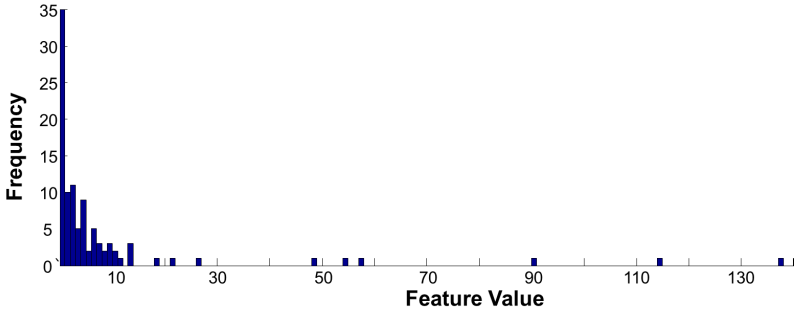

Fig. 5. Feature histogram for adaptive thresholding.

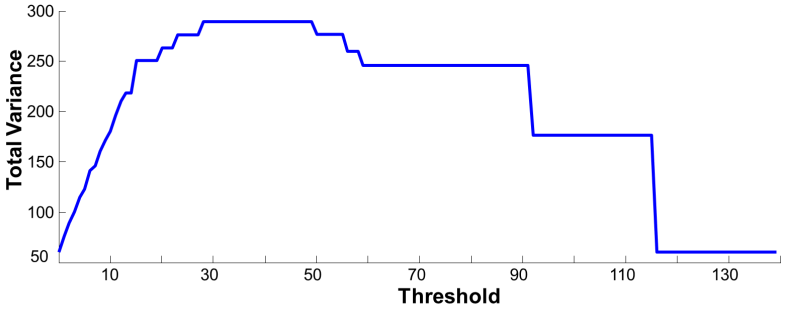

Fig. 6. Relation between threshold and σ_B^2.

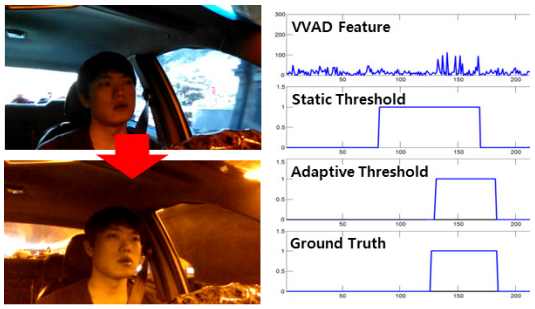

Fig. 7. VVAD result under illumination change situation.

$$\sigma^2 = \sum_{i=1}^{L} (i-\mu_T)^2 p(i), \tag{10}$$

$$\begin{aligned} \sigma_\omega^2 &= \omega_1\sigma_1^2 + \omega_2\sigma_2^2, \\ \sigma_B^2 &= \omega_1(\mu_1-\mu_T)^2 + \omega_2(\mu_2-\mu_T)^2, \end{aligned} \tag{11}$$

이 때 클래스 사이의 분산을 최대화 시키는 T^*가 최적의 문턱값이 된다.

$$T^* = \underset{1 \le T < L}{\arg\max}\ \sigma_B^2. \tag{12}$$

Fig. 5는 한 개의 단어를 발화하는 화자를 촬영한 200장의 영상 프레임에서 추출한 특징의 히스토그램의 예를 나타내고, Fig. 6는 문턱값에 따른 전체 분산의 예를 나타낸다. 음성구간이 입술 움직임 및 조도변화에 의해 특징값이 다양하게 변화하는 상황에서 전체 분산을 통한 문턱값 추정 방법을 적용한 결과는 Fig. 7과 같다. 고정 문턱값을 사용한 경우 차량이 터널을 통과하는 과정에서 조도 변화에 의해 검출률 감소 및 오검출이 증가를 확인할 수 있지만 적응적 문턱값 추정 방법을 적용한 결과는 참 값과 유사한 검출 결과를 나타냄을 알 수 있다.

III. 실험 결과 및 분석

제안한 알고리즘의 성능평가를 위해 주행중인 차량에서 발화하는 화자를 촬영한 동영상을 이용하여 실험을 진행하였다. 5개의 실험 동영상은 총 9,734프레임이며(음성 1,947, 비음성 7,787 프레임) 640 × 480 크기 30 fps의 스마트폰 카메라를 사용하여 촬영했다. Fig. 8은 성능평가용 동영상의 예를 나타낸다.

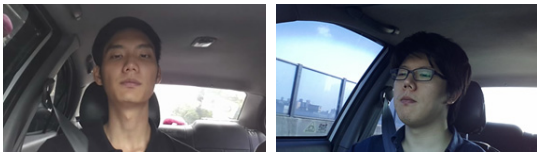

Fig. 8. Example of performance evaluation video.

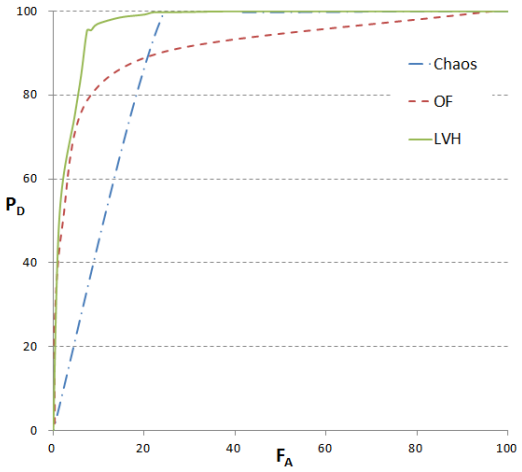

Fig. 9. Experimantal result: ROC curve.

Table 1. Comparison of VVAD algorithm.

		OF[3]	Chaos[6]	LVH[7]	Proposed
DB1	P_D (%)	93.71	70.38	88.43	89.65
	F_A (%)	11.49	24.70	8.53	0.61
DB2	P_D (%)	74.62	97.84	90.96	92.68
	F_A (%)	6.40	10.01	3.12	2.06
DB3	P_D (%)	95.73	97.59	78.04	93.59
	F_A (%)	11.55	7.57	7.19	5.66
DB4	P_D (%)	81.64	37.79	80.27	87.94
	F_A (%)	26.62	4.60	7.22	0.14
DB5	P_D (%)	96.28	89.56	97.97	97.97
	F_A (%)	3.33	4.91	4.20	2.77

각 화자는 약 30개의 짧은 단어에 대해 발화하였다. 성능 측정을 위해 검출률(P_D)과 오검출률(F_A)을 평가지표로 사용하였다. 검출률은 전체 음성구간 프레임의 수와 올바르게 음성구간으로 검출된 프레임 수의 비율로 나타나며, 오검출률은 전체 비음성구간 프레임의 수와 비음성구간인데 음성구간으로 잘못 검출된 프레임의 비율이다. 전체 성능평가 동영상에 대한 ROC 커브는 Fig 9와 같다. 광류 및 카오스 이론 기반 방법을 사용한 결과와 비교하여 지역 분산 히스토그램 기반 특징이 ROC 커브상에서 높은 성능을 나타냄을 알 수 있다. 광류[3] 및 카오스 이론[6]을 사용하는 기존 음성구간 검출 알고리즘과 성능 비교를 수행한 결과는 Table 1과 같다. 광류기반 방법의 경우 높은 검출률을 나타내었지만 상대적으로 높은 오검출률을 나타냄을 알 수 있으며, 카오스 기반 방법은 테스트 동영상에 따라 검출률의 차이가 큰 것을 알 수 있다. 이는 입술 검출 과정에서 발생하는 입술 영상의 평행이동으로 인한 특징값의 변화로 인해 성능의 차이가 나타나는 것으로 고정 문턱값을 사용한 지역 분산 히스토그램 기반 방법[7]의 경우 기존 알고리즘에 비해 오검출률 낮은 오검출률을 나타내었고 검출률 측면에서 낮은 수치를 나타냈으나, 제안한 적응적 문턱값 추정 방법을 사용한 경우 대부분의 평가용 동영상에서 높은 검출률과 낮은 오검출률을 나타냄을 알 수 있다.

IV. 결 론

본 논문에서는 영상기반 음성구간 검출을 위한 지역 분산 히스토그램 특징 및 적응적 문턱값 설정 방법을 제안하였다. 화소의 구조적 패턴 및 Adaboost 알고리즘을 이용하여 조명변화에 강인한 얼굴 및 입술 검출을 수행한 후 지역 분산 히스토그램 기반 특징 추출과 적응적 문턱값 추정 알고리즘을 이용하여 음성구간을 검출하였다. 실험을 통해 기존의 방법인 광류기반 방법 및 카오스 이론기반 방법이 가지는 조도변화 상황 및 연산양 문제를 극복하고 높은 검출률과 낮은 오검출률을 보임을 확인하였다.

감사의 글

본 논문은 삼성전자의 지원을 받아 수행한 연구과제 결과 중 일부이다.

References

1. J. Park, W. Kim, D. K. Han, and H. Ko, “Voice activity detection in noisy environments based on double-combined fourier transform and line fitting,” J. The Scientific World

Journal **2014**, 1-11 (2014).
2. S. Lee, J. Park, Y. Lee, and E. Kim, "Speech activity decision with lip movement image signal" (in Korean), J. Acoust. Soc. Kr. **26**, 25-31, (2007).
3. A. J. Aubrey, Y. A. Hicks, and J. A. Chambers, "Visual voice activity detection with optical flow," J. IET Image Processing **4**, 463-472, (2010).
4. P. Tiawongsombat, M. Jeong, J. Yun, B. You, and S. Oh, "Robust visual speakingness detection using bi-level HMM," J. Pattern Recognition **45**, 783-793 (2012).
5. S. Takeuchi, T. Hashiba, S. Tamura, and S. Hayamizu, "Voice activity detection based on fusion of audio and visual information," in Proc. International Conference on Audio-Visual Speech Processing, 151-154 (2009).
6. K. Lee and H. Ko, "Visual voice activity detection using lip motion and direction in vehicle environment" (in Korean), in Proc. IEEK Fall Conference, 646-647 (2013).
7. G. Kim, J. Ryu, and N. Cho, "Voice activity detection using motion and variation of intensity in the mouth region" (in Korean), J. Broadcast Engineering **17**, 519-528 (2012).
8. T. Song, K. Lee, and H. Ko, "Robust visual voice activity detection using chaos theory under illumination varying environment," in Proc. IEEE International Conference on Consumer Electronics, 574-575 (2014).
9. T. Song, K. Lee, and H. Ko, "Visual voice activity detection via chaos based lip motion measure robust under illumination changes," J. IEEE Transactions on Consumer Electronics **60**, 251-257 (2014).
10. K. Lee, T. Song, S. Kim, D. K. Han, and H. Ko, "Robust visual voice activity detection using local variance histogram in vehicular environments," in Proc. IEEE International Conference on Consumer Electronics, 476-477 (2015).
11. E. Zheng, X, Ping, T. Zhang, G, Xiong, "Steganalysis of LSB matching based on local variance histogram," in Proc. IEEE International Conference on Image Processing, 1005-1008 (2010).
12. B Froba, A Ernst, "Face detection with the modified census transform," in Proc. IEEE International Conference on Automatic Face and Gesture Recognition, 91-96, (2004).
13. J. Beh, R. Baran, and H. Ko, "Dual channel based speech enhancement using novelty filter for robust speech recognition in automobile environment," J. IEEE Transaction Consumer Electronics **52**, 583-589 (2006).
14. N. Otsu, "A threshold selection method from gray-level histograms," J. IEEE Transactions on Systems, Man and Cybernetics **9**, 62-66 (1979).

저자 약력

▸ 송 태 엽 (Taeyup Song)

2005년: 대진대학교 전자공학과, 물리학과 학사
2009년 ~ 현재: 고려대학교 바이오마이크로시스템기술 협동과정 석·박사통합과정 재학중

▸ 이 경 선 (Kyungsun Lee)

2008년: 세종대학교 전자공학과 학사
2012년: 고려대학교 전기컴퓨터공학부 석사

▸ 김 성 수 (Sung Soo Kim)

2008년: 고려대학교 전기전자전파공학부 학사
2010년: 서울대학교 전기컴퓨터공학부 석사
2010년 ~ 현재: 삼성전자 선임연구원

▸ 이 재 원 (Jae-Won Lee)

1993년: KAIST 컴퓨터과학부 석사
1999년: KAIST 컴퓨터과학부 박사
1999년 ~ 현재: 삼성전자 수석연구원

▸ 고 한 석 (Hanseok Ko)

1982년: 미국 카네기 멜론 대학교 전기공학 학사
1986년: 미국 메릴랜드 대학교 시스템공학 석사
1988년: 미국 존스 홉킨스 대학교 전기공학 석사
1992년: 미국 카톨릭 대학교 전기공학 박사
1995년 ~ 현재: 고려대학교 전기전자공학과 교수