PDA 환경에서 자동화자 확인의 계산량 개선을 위한 연구

A Study for Complexity Improvement of Automatic Speaker Verification in PDA Environment

서창우, 임영환, 전성채, 장남영
Chang-Woo Seo, Young-Hwan Lim, Sung-Chae Jeon, Nam-Young Jang

요 약

본 논문은 PDA 디바이스에서 개인정보를 보호하기 위한 자동화자확인 시스템을 제안한다. 최근 M-커머스와 같은 모바일 환경을 위한 PDA의 용량이 확장되고 사용이 증가되고 있다. 그러나 너무 많은 계산량 때문에 PDA 디바이스에서 자동화자확인의 실질적인 응용은 여전히 많은 어려움이 존재한다. 본 논문에서는 이러한 문제점을 해결하기 위해서 음성발성 동안 스펙트럼 차감법과 음성 검출과 같은 전처리를 수행함으로써 계산량을 줄일 수 있는 방법을 적용하였다. 또한 빠른 처리 결과를 얻기 위한 은닉마코프모델의 최적 상태 정합과 시퀀스 확률비 테스트를 적용하였다. 전체적인 시스템은 PDA디바이스의 제한된 메모리와 낮은 CPU 속도에 적합하도록 간결하게 구현하였다.

Abstract

In this paper, we propose real time automatic speaker verification (ASV) system to protect personal information on personal digital assistant (PDA) device. Recently, the capacity of PDA has extended and been popular, especially for mobile environment such as mobile commerce (M-commerce). However, there still exist lots of difficulties for practical application of ASV utility to PDA device because it requires too much computational complexity. To solve this problem, we apply the method to relieve the computational burden by performing the preprocessing such as spectral subtraction and speech detection during the speech utterance. Also by applying the hidden Markov model (HMM) optimal state alignment and the sequential probability ratio test (SPRT), we can get much faster processing results. The whole system implementation is simple and compact enough to fit well with PDA device's limited memory and low CPU speed.

Keywords : automatic speaker verification, PDA, spectral subtraction, mel frequency cepstral coefficient, hidden Markov models, sequential probability ratio test

I. 서 론

지난 10년간 유비쿼터스 환경의 디바이스(모바일 폰, PDA, 노트북, 등)의 개발이 활발해지고 다양한 서비스(정보, 게임, 리모컨 기능, 계산기, 등)가 제공되고 있다. 또한 이들 디바이스의 상업적인 성공을 얻기 위한 연결성과 통신 응용은 점점 더 복잡해지는 잠재적인 응용을 위해서는 필수적이다. 결과적으로, 사용자와 이들 응용들 사이의 상호작용은 중요한 연구 영역이 되고 있으며 모바일 디바이스의 능력에 따라서 다양한 인터페이스(키보드, 햅틱, 펜, 음성, 등)를 최적으로 조합할 수 있게 지원하고 있다. 최근 PDA는 시간과 공간에 상관없이 모바일 환경에서 E-mail, 사이버 주식 시장, 온라인 뱅킹과 같은 다양한 무선 인터넷 서비스를 편리하게 이용할 수 있다[1]. 오늘날의 편리성의 측면에서 발성 명령(spoken commands)을 기본으로 하는 PDA의 응용은 쓰기 명령(written commands)을 기본으로 하는 다른 방법들보다 더욱 편리하다[2,3].

그러나 PDA 사용에 있어서 대두되고 있는 중요한 문제들 중 하나는 사칭자(imposter)의 침입으로부터 사용자(user)의 개인 정보를 보호하는 것이다. PDA 응용에 있어서 적절한 안전을 보장하는 보안 기술(security technology)은 아직 제공되지 않고 있으며[4], 따라서 이러한 문제점을 해결하기 위해서 생체인식 기술(얼굴, 음성, 지문, 홍체, 등)이 대두되고 있다. 이들 방법 중에서 사람의 목소리를 이용한 Automatic Speaker Verification(ASV)가 가장 편리하고 자연스러우며, 그리고 값싼 방법으로서 고려할 수 있다. 그러나 모바일 서비스에 ASV 기술을 적용하기에 가장 큰 어려움은 모바일 기기들의 용량 부족(작은 메모리, 낮은 CPU 속도, 등)에 의한 만족스런 서비스를 제공할 수 없다는 것이다[1]. ASV에서의 대부분의 기존 연구

*창원대학교
투고 일자 : 2009. 4. 4 수정완료일자 : 2009. 7. 29
게재확정일자 : 2009. 7. 29

들은 계산량의 제한 없이 PC 환경에서의 실시간 구현에 초점을 두었기 때문에, 모바일 기기에서의 ASV 기술을 적용하기 위해서는 계산량 감소나 처리 속도를 향상시킬 필요성이 제기 되었다.

PDA에서의 제한된 메모리와 CPU 속도 문제를 해결하기 위해서 Yang에 의해서 실시간 사용자 인증(user authentication) 시스템이 제안되었다[5]. 제안된 방법은 끝점 검출을 위해서 적응적 이중 판별치 끝점 검출(adaptive double thresholds endpoint detection: ADED)과 잡음 문제를 해결하기 위한 특징 클리핑(feature clipping: FC) 기법을 적용하였다. 그러나 끝점 검출과 특징추출에 있어서 속도 개선을 위한 단구간 에너지 기반의 linear prediction coding(LPC)를 이용한 방법은 mel frequency cepstral coefficients(MFCC)를 이용하는 방법보다 성능감쇠라는 문제에 직면할 수 있다[6].

본 논문에서는 PDA에서 동작하는 실시간 ASV시스템을 위한 효과적인 알고리즘을 제안하였다. 제안된 방법은 음성발성 동안 음성처리(speech processing)가 끊임없이 일어나도록 다중-스레드를 통한 세 개의 버퍼를 이용하여 음성데이터를 처리하였다. 또한 일반적인 방법에서의 성능을 유지한 상태에서 속도 개선을 위한 주파수 영역에서 잡음 제거, 끝점 검출, 그리고 특징 추출을 수행하였다. 그리고 최적 상태 정합을 위한 좌향-우향 HMM(LRHMM)[7,8]과 결정 로직을 위한 SPRT[9,10]를 적용하였다. 실험 결과는 제안된 알고리즘을 일반적인 방법과 비교했을 때, 비슷한 성능에서 비교적 많은 계산시간을 줄일 수 있었다.

논문은 다음과 같이 구성되어 있다. II장에서는 음성발성 동안 실시간 특징벡터를 추출하기 위한 과정이 소개된다. III장에서는 속도향상을 위해서 적용된 알고리즘을 설명하였다. 그리고 IV장과 V장에서는 실험 결과 및 결론은 기술하였다.

II. 회전 버퍼를 이용한 실시간 음성 처리

음성발성 동안 실시간 음성처리를 위한 전처리 과정은 그림 1과 같이 진행된다. 현재 PC 환경에서 널리 사용되는 음성 처리는 미리 결정된 시간 동안 발성된 전체 음성 데이터를 수신한 후 소프트웨어개발킷(software development kit: SDK)을 이용하는 방법이다. 그러나 이러한 접근은 PDA와 같은 휴대형 기기(portable device)에서는 발성된 전체 데이터를 한 번에 처리하기에 너무 많은 계산시간이 요구되기 때문에 적합하지가 않다. 이런 문제를 해결하기 위해서 음성발성 동안 스레드를 통한 버퍼를 이용하여 음성 데이터를 처리하는 방법을 고려하였다. 본 논문에서는 웨이브 파일 포맷의 초기화를 위한 SetWaveFormat, 디바이스 오픈을 위한 OpenWaveInDevice, 내부 웨이브 데이터 버퍼의 링크를 위한 SetBuffer, 그리고 각 프레임 버퍼를 위한 AfxBeginThread와 같이 제공된 함수들을 이용하였다.

PDA 기기에서 신호를 입력 받은 후 성능향상을 위해서 잡음과 채널에 대해서 고려를 해야만 한다. ASV 시스템의 성능감쇠는 배경 잡음(background noise), 마이크로폰 왜곡(microphone distortion) 그리고 채널 간섭(channel interference)과 같은 다양한 실질적인 환경에서 일어난다. 이런 열악한 환경에서 ASV 시스템의 신뢰할 수 있는 성능을 얻기 위한 잡음 감소는 반드시 필요하다. 잡음 감소를 위하여 가장 널리 사용되는 기법은 스펙트럼 차감법(spectral subtraction)이다[11]. 특히, 스펙트럼 차감법은 음성에 부과된 잡음을 제거하기에 효과적인 알고리즘이다. 잡음이 섞인 스펙트럼 $N(e^{j\omega})$의 평균 $\mu(e^{j\omega})$은 처음 L개 프레임의 무음(nonspeech) 구간에서 얻을 수 있다. 잡음은 다음과 같이 부가적인 배경 잡음에 의해서 오염된 음성 스펙트럼 크기 $|\overline{H}(e^{j\omega})|$로부터 잡음 스펙트럼의 평균을 제거함으로써 작은 계산량을 이용해서 효과적으로 제거할 수 있다. 이를 식으로 나타내면 다음 식 (1)과 같다.

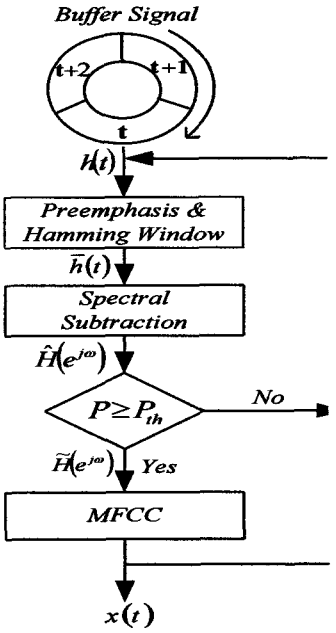


그림 1. 실시간 특징 추출.
Fig. 1. Real-time feature extraction.

$$\hat{H}(e^{j\omega}) = \left[|\overline{H}(e^{j\omega})| - \mu(e^{j\omega})\right] \quad (1)$$

여기서 $\hat{H}(e^{j\omega})$는 잡음이 제거된 추정된 신호(estimated signal)이다.

그림 1과 같이 무음 구간의 잡음 바이어스(noise bias)를 이용하여 효과적으로 잡음을 제거한 후에는 특징 추출(feature extraction)을 위한 음성을 검출해야만 한다. 음성 검출을 위한 방법에서는 에너지(energy)와 시간영역(time domain)에서의 영교차율(zero crossing rate: ZCR)이 널리 사용되지만, 본 논문에서는 식 (2)와 같이 주파수 영역(frequency domain)에서 음성 스펙트럼 $\hat{H}(e^{j\omega})$과 잡음 스펙트럼 $\mu(e^{j\omega})$으로 부터 계산되는 전력비(power ratio) P을 이용하였다.

$$P=20\log_{10}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\hat{H}(e^{j\omega})}{\mu(e^{j\omega})}\right|d\omega\right] \quad (2)$$

음성 검출은 다음과 같이 식 (2)의 전력비 P와 미리 결정된 전력 판별값 P_{Th} 사이의 비(ratio)를 이용해서 얻을 수 있다.

$$\begin{cases}\tilde{H}(e^{j\omega})=\hat{H}(e^{j\omega}) & \text{if } P\geq P_{Th}\\ \tilde{H}(e^{j\omega})=0 & otherwise\end{cases} \quad (3)$$

만약 위의 식 (3)에서 전력비 P가 P_{Th}보다 작다면, 그 프레임은 무음 구간으로 분류된다. 논문에서 사용된 판별값 P_{Th}는 0~2dB이다. 만약 입력 신호가 음성으로 분류된다면 계속해서 다음 과정이 진행되고, 그렇지 않다면 다음 프레임을 위한 새로운 처리가 시작된다. 그리고 전처리 과정에서의 계산 시간을 줄이기 위하여 주파수 영역에서 음성을 검출한 후, 멜 주파수 켑스트럼 계수(MFCC)[12]를 이용하여 직접 특징벡터 $x(t)$을 추출하였다.

Ⅲ. 실시간 ASV 시스템을 위한 화자모델

3.1. 최적 HMM의 상태정합

휴대용 기기에서 ASV 시스템을 지원하기 위해서는 HMM 정합을 위한 빠른 알고리즘이 실시간 구현을 위해서 필요하다[7,8]. 본 논문에서는 속도 문제를 해결하기 위한 검색 공간(search-space)을 줄이기 위해서 비학습적 접근(non-heuristic approach)을 적용하였다. Left-Right hidden Markov model(LRHMM)의 정의에 따라 검색을 위한 부분 공간을 포함하는 변경점(change- points)을 사용하는 순향-후향 기법(forward-backward scheme)에서 HMM 상태들 사이에서 가능한 변경점을 선택하는 것이다. 제안된 방법은 LRHMM에 현재 상태 s에서 다음 상태 $s+1$의 변경점, 그리고 마지막 변경점을 얻기 위해서 반복된다.

시간 t에서 x_t을 d-차수의 특징벡터(feature vector)의 관측 열이라고 하자. 음성처리에서 이것은 정지 구간(stationary)들의 집합을 통해 비정지 과정을 나타내는 상태 열(sequence of state)을 갖는 Markov chain이다. 각 상태에서 음성 데이터의 확률밀도함수(pdfs)는 Gaussian mixture에 의해서 모델링 된다. HMM은 다음과 같이 상태-천이 확률 $A=\{a_{i,j}\}$, 관측 밀도 $B=\{b_j\}$, 그리고 초기 상태 확률 $\Pi=\{\pi_i\}$에 의해 완전히 특성화된다.

$$\lambda=\{A,B,\Pi\}=\{a_{i,j},b_j,\pi_i;i,j=1,2,\cdots,S\} \quad (4)$$

여기서 S는 전체 상태 개수이다. 그리고 관측 벡터 x_t가 주어질 때, 상태 j에서 연속 관측 밀도(continuous observation density) b_j는 다음과 같이 나타낼 수 있다.

$$b_j(x_t)=\sum_{m=1}^{M}w_{jm}N(x_t,\mu_{jm},\Sigma_{jm}) \quad (5)$$

$$\sum_{m=1}^{M}w_{jm}=1 \quad (6)$$

여기서 M은 가우시안 성분 $N(\cdot)$의 전체 개수이며, w_{jm}, μ_{jm} 그리고 Σ_{jm}는 상태 j에서 m-번째 혼합(mixture)의 가중치 함수(weight function), 평균벡터(mean vector) 그리고 공분산행렬(covariance matrix)을 각각 나타낸다. 그리고 상태들 사이의 변경점 검출은 두 데이터 분포들 사이의 변경점을 검출하는 것과 비슷하다. 이것은 LRHMM에서 상태 $s-1$과 s사이의 마지막 변경점을 얻을 때까지 다음 과정을 반복함으로써 구현될 수 있다. 음성 데이터가 시간 t에서 순차적으로 관측될 때, 상태 s에서 $s+1$의 변경(change)은 다음과 같이 일어난다고 가정할 수 있다.

$$R(x_t)=\log\frac{b_{s+1}(x_t)}{b_s(x_t)} \quad (7)$$

여기서 $R(x_t)$는 상태들의 로그 관측밀도 비(ratio)이고 f_s는 다음과 같이 연속관측 개수이다.

$$\tilde{f}_s=\begin{cases}f_s+1 & R(x_t)<0\\ 0 & otherwise\end{cases} \quad (8)$$

지속시간을 나타내는 시간 한계 c_T가 주어질 때, 만약 $f_s\geq c_T$이면, 변경은 상태 s에서 $s+1$로 일어난다. 여기서 작은 c_T는 낮은 false reject(FR)를 갖는 신뢰할 수 있는 변경점을 얻을 수 있지만, false accept(FA)가 증가할 수 있다[7]. 본 논문에서는 실험적으로 신뢰할 수 있는 성능을 얻기 위해서 $c_T=2$을 선택하였다.

위의 알고리즘은 정합에서 어떠한 상태 건너뜀(skip)도 없다는 가정 하에 수행하였다. 그러나 식 (9)와 같이 상태 건너뜀에 대응하는 변경점의 결과로서 고려할 수 있다.

$$R(x_t)=\frac{b_{s+k}(x_i)}{b_s(x_i)} \quad (9)$$

여기서 k는 상태의 최대 건너뜀 수이다. 본 논문에서는 이러한 접근으로 단지 현재 상태는 전체 프레임 대신 각 상태에 속한 프레임만을 계산하기 때문에 계산량을 줄일 수 있다.

3.2. 화자 확인에서 화자와 배경 모델 정규화

본 논문에서는 인증 과정에서 ASV를 위해 배경화자모델(background speaker model)을 사용하였다[13]. 논문에서 적용된 배경화자모델을 이용한 전체 시스템은 그림 2와 같다. 현재 널리 사용되는 배경화자모델에는 여러 가지 방법들이 있지만 universal background models (UBM) 또는 world model이 가장 활발히 사용되고 있다[13]. 그림의 등록 과정에서는 화자의 입력 음성을 이용하여 파라미터 λ_s를 계산하고 그리고 인증 과정에서는 등록 과정에서 구해진 화자 파라미터 λ_s와 많은 사람들의 입력 데이터에 의해서 미리 계산된 배경모델의 파라미터 λ_b를 이용해서 유사도 비를 계산하게 된다. 이때 배경화자모델의 유사도 비(likelihood ratio)는 다음과 같이 나타낼 수 있다.

$$\begin{aligned}l_m(X;\lambda_s,\lambda_b)&=l_s(X,\lambda_s)-l_b(X,\lambda_b)\\&=\frac{1}{T}log\,p(X|\lambda_s)-\frac{1}{T}log\,p(X|\lambda_b))\end{aligned} \quad (10)$$

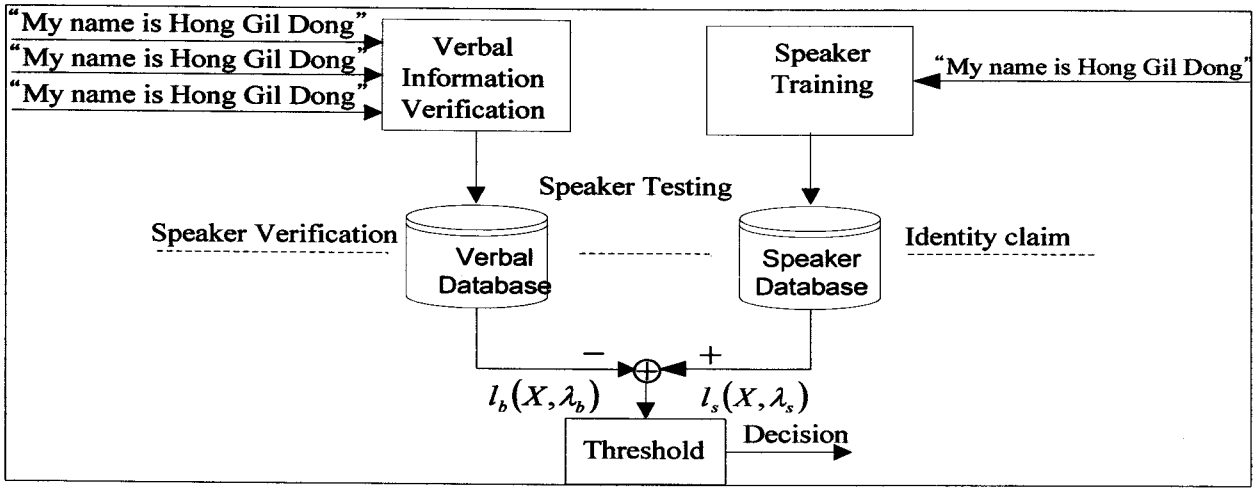


그림 2. 고정 어휘 자동화자 확인 시스템의 구성도.
Fig. 2. An overview of fixed-phrase ASV system.

여기서 $l_s(X,\lambda_s)$와 $l_b(X,\lambda_b)$는 각각 화자와 배경 모델의 유사도이다. $p(X|\lambda_s)$와 $p(X|\lambda_b)$는 화자와 배경 모델로부터 계산된 유사도이다. 인증(acceptance)과 거절(rejection)에 대한 결정은 l_m을 판별값 Th_A와 Th_R에 비교해서 결정할 수 있다.

3.3. 결정로직을 위한 sequential probability ratio test(SPRT)

본 논문에서는 빠른 결정 판별을 얻기 위한 통계적인 시퀀스 결정 방법(statistical sequential decision method)인 SPRT를 적용하였다. ASV 시스템에서 SPRT는 요구되는 입력 샘플을 줄일 수 있기 때문에 계산량 감소를 위한 효과적인 접근 방법이다[9,10]. 여기서 $p_s(x_t)$를 발성 화자(claimed speaker)의 확률밀도함수라 할 때, 결정 로직을 위한 임의의 프레임 F, $1 \le F \le T$에서 SPRT를 위한 유사도 비(likelihood ratio)는 다음과 같다.

$$l_m(x_1,\cdots,x_F)=\prod_{t=1}^{F}p_s(x_t) \tag{11}$$

유사도 비를 이용한 결정 판별은 다음과 같다.

$$l_m(x_1,\cdots,x_F) > Th_A \text{ : Accept as true speaker} \tag{12}$$

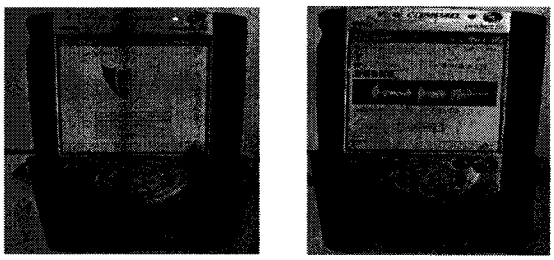
$$l_m(x_1,\cdots,x_F) < Th_R \text{ : Reject as impostor} \tag{13}$$

여기서 Th_A와 Th_R는 $Th_A > Th_R$ 판별값이다. 만약 프레임 F에서 유사도 비 $l_m(x_1,\cdots,x_F)$가 Th_A와 Th_R 사이에 속한다면 결정 판별을 위해서 또 다른 특징벡터를 가져와서 마지막 T까지 반복해서 수행한다. 그리고 인증 또는 거절을 위한 최종 판별은 유사도 비 l_m을 임계 한계값 Th_l에 적용해서 결정을 내려야 한다.

Ⅳ. 시스템 구현 및 실험결과

실시간 ASV 시스템을 구현하기 위해 PDA에서 사용한 O/S는 윈도우 CE 5.0을 사용하였고 개발 툴은 임베디드 비주얼 C++4.0이다. ASV 시스템은 그림 3과 같이 초기화면(initial screen), 등록/인증과정(enrollment/authentication process)로 구성되어 있으며, 편리성을 위해서 패스워드의 입력 없이 단지 식별 번호(identification number: ID)만 등록해서 사용할 수 있도록 구성하였다. 또한 시스템에서의 발성시간은 최대 3초이고 발성시도는 실시간 음성 검출을 적용해서 발성시간의 66%(약 2초)에 도착할 때까지 첫 번째 음성이 입력되지 않으면 자동적으로 종료를 시키고 다시 시도하게 하였다.

실험을 위해 사용된 데이터는 한국어 문장 종속 연속음(text-dependent continuous-speech)을 위해서 "나는 홍길동입니다"를 사용하였다. 데이터 수집에 참가한 화자 인원수는 100명으로 남자 53명과 여자 47명이다. 수집된 데이터에서 남녀 각 25명은 UBM을 위해 사용하였으며, 나머지



(a) (b)

그림 3. PDA 디바이스에서 구현된 ASV 시스템.
(a) 초기 화면, (b) 등록/인증 과정.
Fig. 3. ASV system implemented in PDA device
(a) Initial Screen, (b) Enrollment/Authentication Process.

에서 남녀 각 25명은 UBM을 위해 사용하였으며, 나머지 50명에 대해서 개인별 수집된 8개의 데이터 중에서 등록을 위해 사용한 데이터는 처음 4번 발성한 데이터를 사용하였고 인증 과정은 나머지 4번 발성한 데이터를 사용하였다. 따라서 false reject rate(FRR)을 위해서는 200(50x4)개, false accept rate(FAR)을 위해서는 9800(50x4x49)개의 데이터를 사용하였다. 또한 시스템 구성에 사용된 샘플링 주

파수(sampling frequency)는 11.025kHz이고 분해능은 16bit 이다. 먼저 실험에서 사용된 음성의 프레임 길이는 180 샘플을 한 프레임으로 간주하고 그리고 프레임간 50% 중첩을 적용하였다. 특징벡터는 24개의 필터뱅크(filter bank)로 구성된 12차 MFCC[12]로 구성되어 있으며, 채널의 영향을 줄이기 위해서 cepstral mean subtraction(CMS)[14]를 적용하였다.

본 논문에서 비교 실험을 위해서 사용된 HMM의 상태(state)는 12개, 그리고 각 상태별 3개의 가우시안 혼합(Gaussian mixtures)을 사용하였다. 제안된 방법의 우수성을 확인하기 위해서 전체 검색 알고리즘(full-search algorithm)[15]을 비교 대상으로 실험하였다.

표 1은 인증 과정에서 HMM의 순향 확률(forward probability)를 제안된 방법과 전체 검색 알고리즘과의 계산량 비교를 나타낸 것이다. 표에서 알 수 있듯이 전체 검색 알고리즘은 $X(S^2T)$를 계산하지만 제안된 방법은 $X(T)$만을 계산하기 때문에 훨씬 계산량을 줄일 수 있다.

표 1. 순방향 HMM 알고리즘에서의 계산량 비교.
Table 1. Comparison of computation in forward algorithm of HMM.

Algorithms	Proposed	Full Search
Computation	$X(T)$	$X(S^2T)$

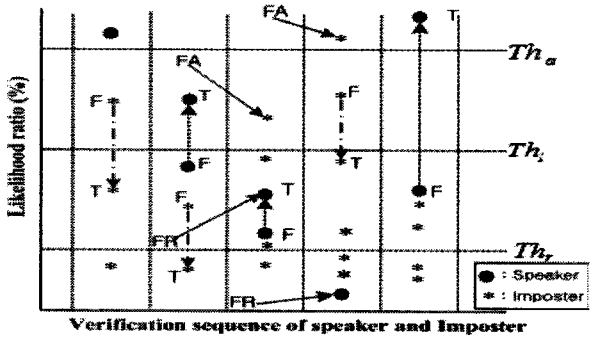


그림 4. SPRT를 이용한 결정 로직 과정.
Fig. 4. Decision logic procedure using SPRT.

그림 4는 ASV에 SPRT를 적용했을 때, 계산시간 및 오류를 줄이는 과정을 설명하고 있다. 여기서 ●와 *는 각각 화자와 사칭자의 유사도를 나타낸다. 만약 프레임 F에서 첫 번째 결정 판별이 $Th_r < l_m < Th_a$에 속한다면, 또 다른 특징벡터가 선택되고 그리고 마지막 프레임 T에서 결정 판별을 하였다. 그림에서 알 수 있듯이 화자(●)인 경우 프레임 F에서 첫 번째 판별값이 Th_l 아래에 있었지만, 프레임 T에서는 Th_l 보다 높은 유사도를 얻을 수 있기 때문에 FR을 줄일 수 있었다. 또한 사칭자(*)인 경우에도 첫 번째 판별에서 Th_l보다 높은 유사도를 얻었지만, 마지막 판별에서 Th_l보다 낮은 판별값을 얻기 때문에 FA를 줄일 수 있다.

표 2는 전체 시스템의 개인별 등록과 인증 과정에서 사용된 평균 계산 시간을 나타낸 것이다. 표에서 알 수 있듯이 등록 과정에서는 효과적인 전처리 과정에 의한 평균 0.16초를 줄일 수 있었다. 또한 인증 과정에서는 전처리 과정과 HMM 상태 정합 그리고 SPRT에 의한 평균 0.023초 시간을 줄일 수 있었다.

표 2. 전체적인 시스템의 평균 계산 시간 비교(s).
Table 2. Comparison of average computational times of total system(s).

Algorithms(Sec)	Proposed (no skip)	Proposed (skip)	Full Search
Enrollment	1.09	1.09	1.25
Authentication	0.0594	0.0587	0.082

표3은 자동화자확인의 성능 평가에 기준이 되는 equal error rates(EER)의 성능을 나타낸 것이다. EER은 확인 성능이 FAR과 FRR이 같게 될 때를 나타낸다. 제안된 알고리즘이 빠른 결정을 얻기 위한 HMM 상태 정합과 SPRT에 의한 감소된 계산량에 의해서 성능이 감소하였지만, 건너뜀(skip)을 적용하지 않았을 때는 전체 검색 알고리즘과 거의 비슷한 성능을 보였다.

Algorithms	Proposed (no skip)	Proposed (skip)	Full Search
EER	3.532	3.627	3.482

표 3. Equal error rates (%)에서의 실험 결과
Table 3. Experimental results in equal error rates (%)

본 논문에서 제안된 방법을 PDA 디바이스에 적용했을 때, 일반적인 방법보다 등록 과정에서는 약 13.7% 그리고 인증 과정에서는 약 25.7%의 계산 시간을 줄일 수 있었다. 이렇게 등록 과정보다 인증 과정의 계산시간이 줄어든 것은 등록과정에서 Ⅱ장의 전처리 과정만 적용되지만, 인증 과정에서는 전처리 과정, HMM 상태 정합, 그리고 SPRT가 함께 적용되기 때문에 훨씬 더 많은 계산시간을 줄일 수 있었다.

Ⅴ. 결 론

본 논문에서는 모바일 환경에서 PDA 기기에 개인 정보 보호를 위한 효과적인 실시간 ASV 시스템을 개발하였다. 제안된 방법은 ASV의 전처리 과정에서 다중-스레드에 의한 프레임 버퍼를 이용함으로써 음성 발성 동안 순차적으로 수행할 수 있었다. 또한 빠른 결정판별을 얻기 위한 HMM에서 상태 정합과 SPRT를 적용하였다. 실험결과 ASV 시스템에서 전체 검색 알고리즘과 비교했을 때, 거의 비슷한 인증율이 나왔지만 계산시간에서 훨씬 빠른 결과로 PDA 디바이스에 적합한 알고리즘인 것을 알 수 있다. 구현된 전체적인 시스템은 제한된 메모리와 낮은 CPU 속도의 PDA 디바이스에 적합하도록 간단하면서도 간결하게 구

성되어 있다.

참 고 문 헌

[1] X. Huang, A. Acero, C. Chelba, etc., "MiPad : a Multimodel Interaction Prototype", *Proc. of ICASSP2001*, pp. 9-12, 2001.

[2] R. Isotani, K. Yamabana, S. Ando, etc., "An Automatic Speech Translation System on PDAs for Travel Conversation", *Proc. of ICMI'02*, 2002.

[3] D. Yong-gang, X. Bo, and H. Tai-yi, "Speech Recognition Algorithm Design and Implementation on Palm PC", *Journal of Computer Research & Development*, vol. 37, no. 8, pp. 937-941, 2000.

[4] T. Iizuka, A. Lau and T. Suda, "A design of local resource access control for mobile agent in PDA", *IEICE Trans. Commun.* vol. E83-B, No. 11, 2001.

[5] Y. Yang, Z. Wu, Z. Ma, "Real time user authentication system for PDA", *Proc. of the IEEE Int. Conf. on Systems, Man and Cybernetics*, pp. 439-443, 2005.

[6] P.C. Joseph, "Speaker Recognition: A Tutorial", *Proc. of the IEEE*, vol. 85, no. 9, pp. 1437-1462, 1997.

[7] Q. Li, "A Detection approach to Search-Space Reduction for HMM Sate Alignment in Speaker Verification", *IEEE Trans. SAP*, vol. 9, no. 5, pp. 569-578, 2001.

[8] Q. Li, "A fast decoding algorithm based on sequential detection of the changes in distribution", *Proc. Int. Conf. Spoken Language Processing*, 1998.

[9] M.A. Lund and C.C. Lee, "A robust sequential test for text-independent speaker verification", *J. Acoust. Soc. Am.*, vol. 99, no. 1, pp. 609-621, 1996.

[10] H. Noda, K. Harada, E. Kawaguchi, "A context-dependent Sequential decision for speaker verification", *IEICE Trans. INF. & syst.* vol. E82-D, no. 10 , pp. 1433-1436, 1999.

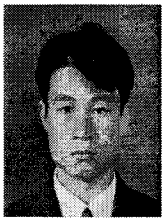
[11] S. F. Boll, "Suppressing of acoustic noise in speech using spectral subtraction", *IEEE Trans. ASSP*, vol. 27, no. 2, pp. 113-119, 1979.

[12] S. Furui, "Cepstral analysis technique for automatic speaker verification", *IEEE Trans. On Acoust, Speech, Signal Processing*, pp. 254-272, vol. 10, no. 4, 1981.

[13] S. Parthasarathy and A. E. Rosenberg, "General phrase speaker verification using sub-word background models and likelihood-ratio scoring", *ICSLP*, vol. 4 ,pp. 2403 - 2406, 1996.

[14] D. Naik, "Pole-filtered cepstral mean subtraction", in *Proc. ICASSP-1995*, vol. 1, pp. 157-160, 1995.

[15] J. K. Chen and F. K. Soong, "An n-best candidates-based discriminative training for speech-recognition applications", *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 206-216, 1994.

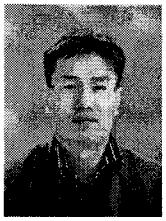


서 창 우(Chang-Woo Seo)

1996년 창원대학교 전자공학과(공학사)
1998년 창원대학교 전기전자제어공학부 (공학석사)
2003년 숭실대학교 정보통신전자공학 (공학박사)
2003년 ~ 2005년 (주)인스모바일 책임연구원
2005년 ~ 2007년 (주)에스씨디 정보통신연구소 책임연구원
2008년 ~ 현재 숭실대학교 IT대학 글로벌 미디어학부 연구교수
※주관심분야 : 음성/영상 신호처리, 모바일 시스템, 멀티미디어 시스템

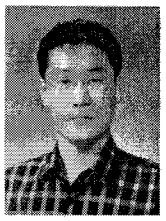
임 영 환(Young-Hwan Lim)

1977년 경북대학교 수학과(이학사)
1979년 한국과학기술원 전산학과(이학석사)
1985년 Northwestern University(이학박사)
1979년 ~ 1996년 한국전자통신연구소 책임연구원
1996년 ~ 현재 숭실대학교 IT대학 글로벌 미디어학부 교수
※주관심분야 : 멀티미디어 공학, 모바일 시스템



전 성 채(Sung-Chae Jeon)

1995년 창원대학교 전자공학과(공학사)
1997년 창원대학교 전자공학과(공학석사)
2006년 한국과학기술원 원자력공학과 (공학박사)
2006년 ~ 현재 한국전기연구원 선임연구원
※주관심분야 : 의공학 영상처리, 생체센서 설계



장 남 영(Nam-Young Jang)

1998년 창원대학교 전자공학과(공학사)
2001년 창원대학교 전자공학과(공학석사)
2006년 창원대학교 전자공학과(공학박사)
2006 ~ 현재 창원대학교 메카트로닉스공학부 연구교수
※주관심분야 : 광 신호처리, 광섬유 센서, 광 계측, 홀로그래피