강인한 주성분 분석법을 갖는 화자인식

이 윤 정, 이 기 용

숭실대학교 정보통신전자공학부

Speaker Recognition Based on Robust PCA

Youn Jeong Lee, Ki Yong Lee

School of Electronic Engineering, Soongsil University,
yjlee@ctsp.ssu.ac.kr, kylee@ssu.ac.kr

요약

본 논문에서는 화자인식을 위하여 강인한 주성분 분석법(Robust Principal Component Analysis)을 갖는 화자인식 방법을 제안하였다. 강인한 주성분 분석법은 특징벡터들의 outlier가 존재할 경우 k-차원으로 줄이면서 강인한 화자 모델을 만들기 위하여 사용한다. 기존의 PCA 방법은 순수한 화자의 정보가 잡음 등의 outlier에 의해 손상될 수 있으므로, 강인한 주성분 분석법을 사용하여 outlier의 영향을 감소 시켰다. 화자 별로 k-차원 diagonal GMM 학습시 mixture 수를 적응시켜 데이터 저장 공간을 최소화하였다. 200명의 고립 숫자음을 사용하여 기존의 diagonal GMM 방법과 제안된 방법을 실험한 결과, 제안된 방법에서 약 1.5% 더 높은 인증률을 얻을 수 있었다.

I. 서론

최근 휴대용 통신기술이 발전함에 따라, 셀룰러 폰과 PDA 같이 휴대용통신기기에 음성을 이용한 다양한 서비스가 제공되고 있지만, 고속의 PC환경에서 실험되었던 음성 분야들을 휴대용 기기에 적용하기 위해서는 계산량과 처리속도 등의 문제를 해결 해야한다.

GMM(Gaussian Mixture Model)은 speaker verification과 identification을 위해 많이 사용되고 있고[1], PCA(Principle Component Analysis)는 공분산(covariance)의 고유값과 고유벡터를 이용하여 다중 변환 데이터를 분석하기 위하여 많이 사용 되고 있다[2]. 비록 GMM은 diagonal 공분산 행렬을 가지더라도, mixture 수가 클수록, 특징벡터가 많을수록 화자 인식 성능을 향상시킬 수 있다. 그러나 특징벡터의 차원과 mixture 수가 늘어나면, 인식을 위해서 저장공간이 많이 요구될 뿐만 아니라, 계산량이 많아지고 실시간 구현이 어렵게 된다. 그리고, 잡음과 같은 outlier가 특징벡터에 포함될 경우 화자의 순수한 특징을 추출하기 어렵다[3].

따라서, 본 논문에서는 outlier가 포함된 특징 벡터의 차원을 줄인면서, 화자의 순수한 정보를 추출하기 위하여 강인한 주성분 분석법과 각 화자의 기존 diagonal GMM 학습시 화자별로 다른 mixture 수로 적응시키는 방법을 제안하였다.

본 논문의 구성은 다음과 같다. 화자인식을 위하여 II 장에서는 기존의 주성분 분석법과 강인한 주성분 분석법을 소개하였다. III장에서는 강인한 주성분 분석법을 이용한 GMM 방법을 제안하였고, IV장에서는 기본 diagonal GMM 방법과 제안된 방법을 200명의 고립 숫자음을 사용하여 결과를 비교하였다. 마지막으로 결론을 내린다.

II. 강인한 주성분 분석법

II-1. 주성분 분석법 (Principal Component Analysis)

주성분 분석법(PCA)은 상태열 $X=\{\vec{X}_1,\vec{X}_2,\cdots,\vec{X}_T\}$ 인 학습 특징벡터가 여러 개의 관련성이 밀접한 변수들 $\vec{X}_t=[x_1,x_2,\cdots,x_p]$ 로 구성되어 있을 때, 서로 상관관계가 없고, 정보의 손실 없이 그 수가 p 보다 작은 k차원의 새로운 변수 즉, 주성분 벡터인 $\vec{Y}_t=[y_1,y_2,\cdots,y_k]$ 로 축약시키기 위한 다변량 분석 기법으로 선형 결합 함수를 구하는 것이다.

PCA는 $\vec{X}_t=[x_1,x_2,\cdots,x_p]$, $t=1,2,\cdots,T$ 에서 평균 벡터 $\vec{C}$ 와 분산을 가지는 선형 변환 행렬 Ω^T 을 계산하기 위해 σ_{ij} 를 원소로 갖는 $p\times p$ 공분산 행렬 Σ_X 를 구해야 한다. 공분산행렬 Σ_X 은 고유 값과 고유벡터로 나누어진다.

$$\Sigma_X = \sum_{i=1}^{p} \lambda_i \upsilon_i \upsilon_i^T \quad (1)$$

여기에서 λ_i 는 Σ_X 의 i번째 고유값이고, υ_i 는고유값 λ_i , $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p)$ $i=1,2,\cdots,p$ 이 주어졌을 때 정 대응되는 정규화된 고유벡터이고, 이들은 $p \times p$ 인 직교 행렬($\Omega\Omega^T = I$)을 이룬다. 이로부터, t 번째 상태열의 특 징벡터 $\vec{X}_t$ 와 주성분 벡터 $\vec{Y}_t$ 의 관계는

$$\vec{Y}_t = \Omega^T \vec{X}_t \quad (2)$$

로 나타낼 수 있다. 여기에서 Ω^T 는 X 를 Y 로 선형 변환하기 위한 크기가 $p \times p$ 인 변환 행렬이다.

만약, k=p이면, 위의 식은 $\vec{X}_t$ 에서 $\vec{Y}$ 으로 손실없이 원자료의 전체 공분산이 100% 표현됨을 알 수 있다. 고유값의 크기순 $(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq \cdots \geq \lambda_p \ (k < p))$ 으로 나타낼 때, λ_i 의 짝을 υ_i 으로 나타내어 k -차원에 해 당되는 Ω_S 을 선택한다.

$$\Sigma_Y = \sum_{i=1}^{k} \lambda_i \upsilon_i \upsilon_i^T, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \quad (3)$$

$$\Omega_s = [\upsilon_1 \ \ \upsilon_2 \ \ \upsilon_3 \ \ \cdots \ \ \upsilon_k] \quad (4)$$

k -차원 주성분 벡터의 정보 비율(I)은 다음식에 의해 구할 수 있다.

$$I = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \quad (5)$$

이 정보비율에 의해 고유값이 큰 것부터 k -차원 만 선택하여 $\Omega_S{}^T$ 를 구하고,

$$\vec{Y}_t = \Omega_S{}^T \vec{X}_t \quad (6)$$

로 주성분 벡터를 얻을 수 있다.

II-2. 강인한 주성분 분석법 (Robust PCA)

II-1에서 설명한 전형적인 PCA 방법은 주변의 outlier가 포함된 특징벡터들의 경우 불안정하다. 따라서, 이를 해결 하기 위하여 차원을 줄이면서 outlier가 포함된 특징벡터를 강인하게 만드는 강인한 주성분 분석법을 제안하였다.

i번째 성분의 평균 벡터 $\vec{C}_i$ 와 분산 벡터 $\vec{V}_i$ 를 이용하여 특징벡터와 평균 벡터 사이의 거리 d_i 를 측정 한다.

$$d_i = \sum_{i=1}^{p} \frac{(x_{t,i} - \vec{C}_i)^2}{V_i} \quad (7)$$

만약 특징벡터와 평균 벡터 사이의 거리가 경계값 q_δ 보다 크다면, Huber weight 함수[5]

$$\begin{cases} w_t = 1 & d_t < q_\delta \\ w_t = \dfrac{q_\delta \, \mathrm{sgn}(d_t)}{d_t} & d_t > q_\delta \end{cases} \quad (8)$$

를 적용하여 잡음과 같은 outlier에 강인한 평균과 분산

$$\hat{\vec{C}} = \frac{\sum_{t=1}^{T} w_t \vec{X}_t}{\sum_{t=1}^{T} w_t} \quad (9)$$

$$\hat{\mathrm{V}} = \frac{\sum_{t=1}^{\mathrm{T}} \mathrm{w}_t (\bar{\mathrm{X}}_t - \hat{\bar{\mathrm{C}}})(\bar{\mathrm{X}}_t - \hat{\bar{\mathrm{C}}})^{\mathrm{T}}}{\sum_{t=1}^{\mathrm{T}} \mathrm{w}_t} = \Sigma_{\mathrm{X}} \quad (10)$$

을 재계산 한다.

식(10)로부터 변환행렬 $\hat{\Omega}_S^T$ 를 구하여, t 번째 상태열 의 특징벡터 $\vec{X}_t$ 와 주성분 벡터 $\vec{Y}_t$ 의 관계를 식(11)로 표현한다.

$$\vec{Y}_t = \hat{\Omega}_S^T \vec{X}_t \quad (11)$$

III. 강인한 주성분 분석법을 이용한 GMM

III-1. GMM 학습과정

(a) Robust PCA를 이용한 GMM

p-차원을 가지는 상태열 T개의 학습 벡터를 $X = \{\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_T\}$ 라 두자. II장에서 제안한 강인한 주 성분 분석법을 이용하여 주성분 벡터의 차원을 k 로 정한 다. 이를 사용하여 가우시안 성분 밀도 함수는 성분의 가 중치(weight), 평균벡터(mean vector), 분산 행렬(variance matrix)로 나타 낼 수 있다.

$$\theta = \{p_i, \vec{\mu}_{Yi}, \Sigma_{Yi}\}, \quad i = 1,...,M \quad (12)$$

$k \times p$ 변환 행렬을 사용하여 식(11)에 의해 T개의 k -차원 학습 주성분 벡터 $\vec{Y}_t = [y_1, y_2, \cdots, y_k]$ 를 구하였다. 변 환된 k -차원 주성분 벡터 Y를 이용한 GMM의 유사도는

$$p(Y \mid \lambda) = \prod_{t=1}^{T} p(\vec{Y}_t \mid \theta) \quad (13)$$

로 구할 수 있다[1]. $p(\vec{Y}_t \mid \theta)$ 는 M성분의 확률밀도값 의 가중된 합이다.

$$p(\vec{Y}_t \mid \theta) = \sum_{i=1}^{M} p_i b_i(\vec{Y}_t) \quad (14)$$

여기서, $b_i(\vec{Y}_t)$ 는 k -차원 가우시안 성분 밀도값이고, 각각의 성분 밀도값은

$$b_i(\vec{Y}_t) = \frac{1}{(2\pi)^{k/2} |\Sigma_{Yi}|^{1/2}} \exp\left[-\frac{1}{2} \left\{ (\vec{Y}_t - \vec{\mu}_{Yi}) \Sigma_{Yi}^{-1} (\vec{Y}_t - \vec{\mu}_{Yi}) \right\} \right] \quad (15)$$

평균 벡터 $\vec{\mu}_{Yi}$ 와, 공분산 행렬 Σ_{Yi} 로 나타 낼수 있다.

GMM의 확률값을 최대로 하기 위해 ML(Maximum Likelihood)알고리즘을 사용하여 모델 파라미터 (θ) 를 찾 는다. ML을 이용하여 파라미터는 EM(Expectation-maximization)을 이용하여 반복적으로 계산하여 구할수 있다. 재추정된 가중치, 평균벡터, 분산 행렬은 다음 식 으로 표현 할 수 있다.

- 성분의 가중치(mixture weight)

$$\bar{p}_i = \frac{1}{T}\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta) \tag{16}$$

- 평균 벡터(mean vector)

$$\hat{\vec{\mu}}_i = \frac{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)\vec{Y}_t}{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)} \tag{17}$$

- 분산 행렬(variance matrix):

$$\hat{\vec{\sigma}}^2_i = \frac{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)\vec{Y}_t^2}{\sum_{t=1}^{T} p(i|\vec{Y}_t,\theta)} - \hat{\vec{\mu}}^2_i \tag{18}$$

- 사후확률(A posterior probability):

$$p(i|\vec{Y}_t,\theta) = \frac{p_i b_i(\vec{Y}_t)}{\sum_{m=1}^{M} p_m b_m(\vec{Y}_t)} \tag{19}$$

(b) GMM을 위한 화자별 mixture 수

본 논문에서는 speaker identification을 위해 화자의 학습 과정에서 T 개 상태열의 주성분 학습 벡터의 k - 차원 GMM 유사도 계산시 M 개의 mixture 수를 초기값으로 두고,

$$\bar{p}_i < \frac{1}{M} \quad \text{이면,} \quad \hat{M} = M - 1 \tag{20}$$

로 가장 작은 가중치값을 갖는 i_{min} 번째 mixture를 제거한 후 mixture의 가중치 $\bar{p}_i$ 를 줄어든 mixture의 수로 정규화 시킨다.

$$\bar{p}_i = \frac{1}{\hat{M}} \tag{21}$$

(c) 학습을 위한 알고리즘 순서도

화자인식을 위하여 robust PCA에 근거한 GMM 알고리즘을 순서도를 그림1에 나타내었다. background DB에서 저장한 평균벡터와 분산벡터를 이용하여 화자의 특징벡터의 outlier 문제를 해결하였다. 변환행렬을 사용하여 PCA를 사용하여 특징벡터의 차원을 줄이고, 화자의 학습 모델의 초기값은 background DB를 사용하였다.

III-2. GMM 테스트 과정

화자의 음성이 입력 되면, 특징벡터 X 를 구한 뒤 background DB의 변환행렬(Ω)을 이용하여 특징벡터를 선형변환하여 주성분 벡터 Y 를 구하였다.

S명 화자 각각은 robust GMM의 $\theta_1,\theta_2,\cdots,\theta_s$ 로 나타내고, 화자의 주성분 벡터를 이용하여 GMM의 최대 사후확률 값을 갖는 화자모델 n 를 찾을 수 있다.

$$\hat{S} = \arg\max_{1\le n\le S} \sum_{t=1}^{T} \log p(\vec{Y}(t)|\theta_n) \tag{22}$$

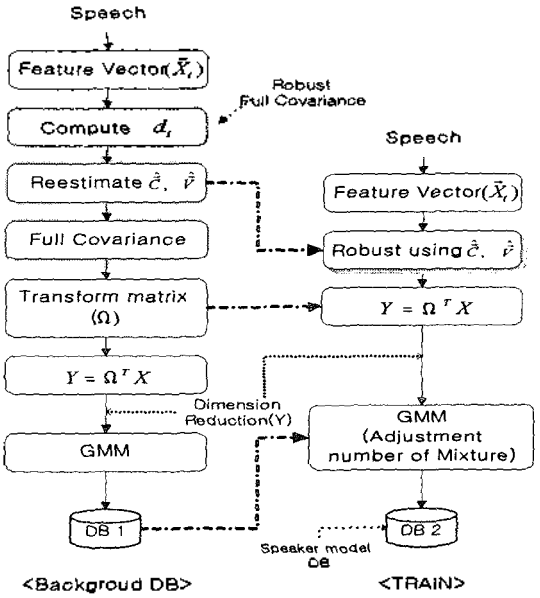

그림1. Background DB와 화자의 GMM 학습 과정

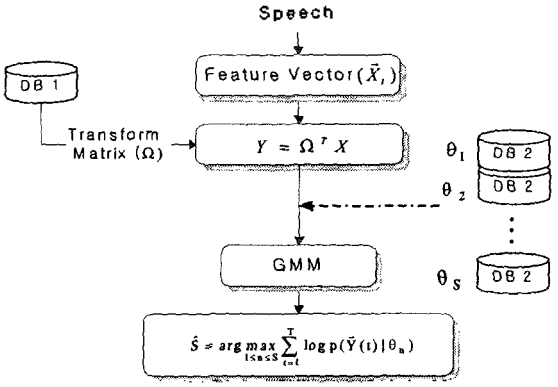

그림2. Speaker identification을 위한 테스트 과정

IV. 실험 및 결과

IV-1. 실험

본 논문에서 제안한 방법을 검증하기 위하여 실험에 사용된 음성은 200명(남자 114명, 여자 86명)의 화자가 10개의 고립 단어 숫자음(영, 일, 이, …, 칠, 팔, 구)을 1회에 4번씩 발음한 뒤, 7일~30일 사이의 시간 간격을 두고 2회에 4번 발음한 데이터이다. 음성 데이터 중에서 실험적으로 성능이 좋은 영, 일, 삼, 칠, 팔을 선택하여 실험에 사용하였다. 11025Hz로 샘플링 하였고, 특징벡터로는 12차원 LPC 켑스트럼과 13차원 delta 켑스트럼을 사용하였다. 음성 분석을 위하여 hamming window가 사용 되었고, 한 프레임은 90샘플 중첩을 가진 180샘플을 사용하였다. 각 화자가 1회에 녹음한 4개의 음성 데이터를 학습에 사용하였고, 2회에 녹음한 4개의 음성 데이터를 테스트에 사용하였다. 강인한 주성분 분석법 사용을 위하여 정보율이 95%이상인 17차원을 선택하였다. 실험을 위

해서 outlier는 특징벡터 프레임의 3~5%에 해당하는 프레임을 임의로 변화시켜 사용하였다.

IV-2. 결과 및 고찰

200명 화자의 음성 데이터로 기존의 diagonal 행렬을 갖는 p-차원 GMM 방법과 화자별로 mixture의 수가 다른 GMM 방법과 제안된 방법을 비교하였다.

표1은 기존 diagonal GMM의 방법과, 제안된 방법의 성능을 나타낸 것이다. 특징벡터에 outlier를 0%, 3%, 5% 추가하여 실험 하였을 때, 제안된 방법이 약 1.5% 정도 우수한 성능을 보였다. 0%, 3%의 outlier를 추가한 경우, 제안된 방법은 약간의 outlier가 존재할 때 깨끗한 음성에서 보다 오히려 성능이 높게 나타났는데, 이는 robust 알고리즘은 약간의 outlier가 존재할 때 더 좋은 성능을 나타냄을 알 수 있었다. 또한, 제안된 방법에서 outlier를 5% 추가한 경우와 깨끗한 음성의 경우를 비교하면 0.3%정도의 차이가 있었다. 또한, 기존의 GMM 방법은 5%의 outlier 추가시 GMM 학습 과정에서 불안정하게 수렴 하여, outlier가 존재할 때 robust한 방법이 필요함을 알 수 있었다.

표2는 전체 200명 화자의 GMM 학습시 걸리는 시간과 DB2에 저장되는 각 화자의 파라메타 수를 나타낸 것이다. 제안된 방법이 특징벡터의 선형변환 과정이 추가 되었지만, 화자의 학습과정에서 GMM 유사도 계산시, EM 알고리즘의 반복 과정에서 차원이 감소되어 기존 방법보다 32초 적게 걸렸다. DB2에 저장되는 화자의 파라메타 수는 M 이 초기 mixture수, $\hat{M}$ 이 조정된 mixture 수, p가 특징벡터의 차원, k가 주성분 벡터의 차원일 때, $\hat{M} \le M$, $k<p$ 이므로 제안된 방법이 기존의 GMM방법보다 적은 저장 공간이 필요하였다.

표1. 기존 diagonal GMM의 화자별 mixture 수를 고정/변화한 방법과 제안된 방법의 성능(%)

방법 \ Outlier 추가		0%	3%	5%
GMM	화자의 mixture 수 고정	91.250[%]	91.125[%]	90.875[%]
	화자의 mixture 수 변화	92.375[%]	91.875[%]	91.375[%]
Robust PCA (제안된 방법)		92.625[%]	92.750[%]	92.375[%]

표2. 기존의 방법과 제안된 방법의 학습시간 및 파라메타 수

방법	학습시간 (sec)	DB2의 파라메타 수
GMM	252	$M+M\times p+M\times p$
Robust PCA (제안된 방법)	220	$\hat{M}+\hat{M}\times k+\hat{M}\times k$

(M :초기 mixture수, $\hat{M}$: 조정된 mixture수 : $\hat{M} \le M$)
(p:특징벡터의 차원, k:주성분 벡터의 차원 : $k<p$)

V. 결론

본 논문에서 제안한 방법은 $k=p$일 때, 기존의 diagonal GMM 방법으로 대체 할수 있다. 화자마다 5개의 고정된 mixture와 $p=25$ 차인 기존의 diagonal GMM 방법과 제안된 방법을 비교할 때, 학습 시간이 12.6%가 감소되었고, $\hat{M} \le M$, $k<p$ 이므로 학습시 파라메타 수를 줄일 수 있었다. 제안된 방법은 깨끗한 음성에서 보다 약간의 outlier가 존재 할 때 더 좋은 인증률을 가졌다. 제안된 방법의 성능이 기존의 diagonal GMM방법보다 약 1.5% 정도 우수하지만, 표1에 나타난 성능은 최근 국내외의 화자 인식 분야에서의 실험 결과와 비교 하였을 때 높은 성능이 아니다. 이는 학습에 사용된 음성과 테스트에 사용된 음성과의 시간차이 때문에 발생되었다. 학습과정에 사용된 음성과 테스트에 사용된 음성을 하나씩 바꾸어 speaker identification 하였을 때 기존 diagonal GMM 방법은 100[%]의 인증률로 나타났다. 따라서, GMM 방법에 시간 흐름에 따른 화자별 적응 과정에 대한 연구가 필요하다.

참고문헌

[1] Reynolds, D.A., and Rose, R.C., "Robust Text- Independent Speaker Identification Using Gaussian Mixture Speaker Models", *IEEE Trans. SAP, vol. 3, No. 1, 1995, pp. 72-83*

[2] Liu, L., and He, J., "On the Use of Orthogonal GMM in Speaker Recognition", *ICASSP, Proc., 1999, pp.845-849*

[3] Changwoo, S., KiYong, L., Joohun, L.,"GMM based on local PCA for Speaker Identification", *Electronics Letters, Vol., 37, No. 24, 22nd 2001, pp.1486-1488.*

[4] Croux, C., and Haesbroeck, G., "Principal Component Analysis Based on Robust Estimators of the covariance or correlation matrix: Influence function and efficiencies", *Biometrika, Vol.87, No.3, 2000, pp.603-618*

[5] P.J. Huber, *Robust Statistics*. New York: Wiley, 1981.