

UTV 분해를 이용한 KLT-CVQ 코더의 계산량 개선

주현호 *김무영

세종대학교

aagusgh@googlemail.com *mooyoung@sejong.ac.kr(교신저자)

Reduction in Computational Complexity of KLT-CVQ using UTV Decomposition

Hyunho Ju *Moo young Kim

Sejong University

요약

사람의 음성을 압축하는 방법으로 Code Excited Linear Prediction (CELP) 코더가 주로 사용되어 왔다. CELP 코더의 수신단에서는 양자화 된 여기신호를 LPC 필터로 합성하여 신호를 복원한다. LPC 합성필터의 영향으로 양자화 된 여기신호의 보로노이 셀 모양이 변형되는 문제점이 있기 때문에 이런 문제점을 해결하기 위해서 Karhunen-Loeve-Transform based Classify Vector Quantization (KLT-CVQ) 코더가 제안되었다. 기존 KLT-CVQ 코더는 KLT 변환과 class 선택을 위해서 Eigen Value Decomposition (EVD)을 이용해서 eigen vector와 eigen value를 계산한다. 본 논문에서는 EVD 대신에 UTV Decomposition (UTVD)을 이용하여 KLT-CVQ의 계산량 문제점을 개선하는 방법을 제안한다.

1. 서론

사람의 음성을 효율적으로 압축 및 전송하는 방법으로 Linear Prediction Coding (LPC) 계수를 이용하는 Code Excited Linear Prediction (CELP) 코더가 주로 이용된다. CELP 코더는 낮은 비트율을 통하여 효율적인 지각 성능을 얻을 수 있다. 하지만 CELP 코더에서 사용되는 LPC 계수는 여기 도메인에서 동작하게 되는데 이때 양자화된 여기신호를 LPC 필터를 이용하여 합성하는 과정에서 보로노이 셀 모양이 변형되는 문제점이 생긴다. 이러한 문제점을 해결하기 위해서 기존신호 값을 여기 도메인이 아닌 Karhunen-Loeve-Transform (KLT)을 이용한 KLT 도메인에서 처리 하는 KLT-Classify Vector Quantization (KLT-CVQ) 코더가 제안 되었다 [1]. 이 방식을 사용하면 CELP 코더의 문제점이 해결되어서 성능을 증가 시킬 수 있다.

KLT-CVQ 코더에서는 입력 신호를 eigen vector를 이용하여 KLT 도메인으로 변환하여 처리를 하고, 입력신호와 에너지 분산이 유사한 코드북의 찾기 위해서 eigen value를 이용하는데 코드북은 에너지 분산별로 class가 나누어져 있으며 이 중에서 왜곡이 적은 코드북의 class를 선택한다. 이 과정에서 eigen value와 eigen vector는 공분산 행렬을 EVD 하여 구한다. 사용되는 Eigen Value Decomposition (EVD)은 Singular Value Decomposition (SVD)과 주로 신호처리에서 행렬을 분해하는 방법으로 많이 사용된다. 하지만 KLT-CVQ 코더에서 EVD를 이용하는 방법은 신호의 차원이 증가할수록 높은 계산량이 요구된다.

본 논문에서는 EVD 대신에 UTV Decomposition (UTVD) [2]을 사용하는 방법을 제안한다. UTVD는 eigen value와 eigen vector를 구하는데 EVD보다 적은 계산량을 요구하기 때문에 KLT-CVQ 코더에 적용한다면 차원의 증가에 따른 계산량 증가 문제를 해결 할 수 있다.

2. 본론

사람의 음성은 과거 신호와 현재 신호 둘 사이에는 변화가 적다는 특성을 갖고 있다. 이러한 음성의 특성이 있기 때문에 CELP 코더는 backward-adaptive 방식을 이용하여 LPC를 구하는 방법으로 사용한다. 그렇기 때문에 송신하는 쪽과 수신하는 쪽에 추가로 전달하는 정보를 줄일 수 있게 되어서 비트율을 줄일 수 있다. CELP 코더에서는 여기신호를 LPC 합성 필터를 이용하여 신호를 합성하게 된다.

$$\hat{s}^p = H\hat{e}^p \quad (1)$$

여기서 \hat{s}^p 는 합성된 음성 신호이고, H 는 LPC 합성 필터, \hat{e}^p 는 양자화된 여기신호를 나타낸다. 한 프레임에 음성을 5샘플을 사용하여 p 가 5인 LPC 합성필터 H 는 다음과 같이 나타낼 수 있다.

$$H = \begin{bmatrix} h(0) & 0 & 0 & 0 & 0 \\ h(1) & h(0) & 0 & 0 & 0 \\ h(2) & h(1) & h(0) & 0 & 0 \\ h(3) & h(2) & h(1) & h(0) & 0 \\ h(4) & h(3) & h(2) & h(1) & h(0) \end{bmatrix} \quad (2)$$

이 합성필터 H 는 LPC의 impulse response를 원소로 갖는다. 그러나 양자화된 여기신호를 LPC 합성필터를 통하여 음성신호로 복원 할 때 합성하는 과정에서 보로노이 셀 모양이 변형되는 문제점을 가지게 된다. 이러한 문제점을 해결하기위해 KLT-CVQ 코더가 제안되었다.

KLT-CVQ 코더에서는 입력 음성의 영상태 응답을 KLT하여 처리한다.

$$y^p = Ux^p \quad (3)$$

x^p 는 입력 신호의 영응답 상태 이고, y^p 는 영상태 응답을 KLT 도메인으로 변환한 값이다. 이때 U 는 eigen vector이고 KLT를 하기 위해 사용한다. U 행렬을 구하는 방법에는 신호의 합성에 이용되는 LPC 합성필터인 H 를 이용하여 구할 수 있다. 기존의 CELP에서 사용되는 H 행렬은 SVD를 사용하여 다음과 같이 분해 할 수 있다.

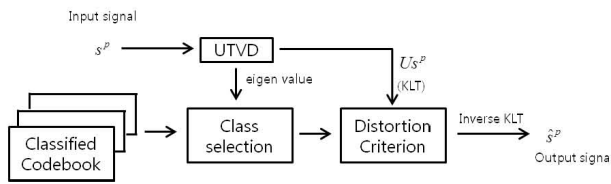


그림 1. UTVD를 이용한 KLT-CVQ 코더의 블록도

$$H = U\Sigma V^T \quad (4)$$

여기서 W 와 V 는 신호의 공분산 행렬인 $H^T H$ 의 eigen vector 행렬이다. 그리고 Σ 는 singular value를 갖는 대각행렬인데, 이때 H 행렬은 정사각행렬이기 때문에 SVD를 EVD로 대체하여 사용할 수 있다. 그러므로 eigen value는 Σ 행렬의 원소들의 제곱이다. 이렇게 구한 eigen vector는 KLT를 할 때 사용되고, eigen value는 KLT-CVQ 코더에서 코드북 class를 선택할 때 사용되게 된다. 하지만 SVD 혹은 EVD를 이용할 때 2가지 문제점을 갖게 된다. 첫 번째는 동작하는데 있어서 많은 계산량을 요구 하고, 두 번째는 null space를 업데이트 하는 과정이 매우 어렵다는 것이다. 이러한 SVD와 EVD 단점을 해결하기 위해서 UTVD가 제안되었다. 이 방법은 분해 결과에서 singular value를 원소로 갖는 대각행렬을 삼각행렬로 표현한다. 이러한 UTVD를 계산하는 방법은 행렬의 0값을 업데이트 하는 평면 회전 방법이 사용된다. 평면 회전 방법을 이용한 UTVD는 SVD의 계산량인 $O(n^3)$ 보다 감소된 계산량 $O(n^2)$ 을 얻을 수 있다. (4)번 식과 비교하여 UTVD가 계산되는 과정은 다음과 같이 나타낼 수 있다.

$$H = UTV^T \quad (5)$$

여기서 T 는 삼각행렬이다. UTVD의 또한 3개의 행렬로 분해되며, 그 중에서 가운데 분해된 삼각행렬의 모양에 따라서 대표적으로 2개로 나뉜다. 행렬이 상삼각행렬 이라면 URV Decomposition (URVD)라 부르고, 하삼각행렬이라면 ULV Decomposition (ULVD)라고 부른다.

여기서 UTVD의 결과로 나온 삼각행렬의 대각선 성분은 eigen value의 근사값을 갖고, 나머지 성분들은 0에 매우 가까운 값을 갖는 성질을 갖고 있다. 그러므로 삼각행렬에서 대각선 성분만을 이용한다면 기존의 KLT-CVQ 코더에 적용을 할 수 있다. 그림 1은 UTVD가 적용된 KLT-CVQ 코더의 블록도를 나타낸다.

3. 실험 및 결과

실험을 하기 위해서 사용한 음성은 8kHz로 다운샘플링된 TIMIT 데이터를 사용하였다. 340개의 음성을 이용하여 학습데이터에 사용하였고, 100개의 음성을 실험데이터로 사용하였다. 기존의 KLT-CVQ의 성능과 비교하기 위해서 한 프레임의 길이는 0.625ms이며, 합성을 위한 LPC 계수는 50자를 사용하고, formant weighting과 gain 예측을 위한 LPC 계수는 10자를 사용하였고, 2.5ms 마다 LPC계수들을 분석하였다. 그리고 class의 수를 1, 2, 4, 8, 16, 32까지 증가 시켜가며 실험을 하였다.

KLT-CVQ 코더에 사용되는 두 개의 코드북인 shape와 gain은 각 각 7비트와 3비트를 주어 비트율은 16kbps이다.

Shape 코드북은 학습데이터의 에너지 분산을 갖는 가우시안 데이터를 이용하여 만든다. 이때 백터 양자화를 이용하여서 코드북을 생성

해준다.

Gain 코드북은 forward 방식을 이용하여 optimal gain을 구하고, 그 값을 양자화 하여 사용하였다.

KLT를 위한 eigen vector와 class 선택을 위한 eigen value를 구하기 위한 방법으로 EVD, UTVD를 사용하여 학습과 실험에 모두 적용 하였다. 실험의 성능은 Signal to Noise Ratio (SNR)을 이용하여서 성능을 비교 하였다. 그에 대한 수식은 다음과 같다.

$$SNR = 10 \log_{10} \frac{s^p}{\hat{s}^p} \quad (6)$$

위 식에서 s^p 는 입력 신호이고, \hat{s}^p 는 KLT-CVQ 코더를 사용하여 결과 값으로 출력된 신호이다.

표1에서는 KLT-CVQ 코더를 EVD와 UTVD를 사용하여 설계한 후, class 수를 증가시키며 SNR 성능을 비교하였다. EVD와 UTVD를 사용한 경우의 SNR 성능을 비교 하였을 때, 두 방식의 성능이 거의 유사한 것을 볼 수 있다. 계산량 관점에서 살펴보면, EVD의 계산량이 $O(n^3)$ 임에 비해서 UTVD의 계산량은 $O(n^2)$ 이다. 따라서, UTVD를 사용하여 KLT-CVQ를 설계하면 EVD를 사용한 경우와 비교하여 훨씬 적은 계산량으로 유사한 수준의 SNR 성능을 보임을 알 수 있다.

표 1. Decomposition 방법에 따른 KLT-CVQ 코더의 SNR 성능

class 수	SNR (dB)	
	EVD	UTVD
1	17.0492	16.9654
2	17.3472	17.4287
4	17.4939	17.5063
8	17.5506	17.5141
16	17.5571	17.5417
32	17.6108	17.5609

4. 결론

최적의 VQ 구현을 통한 기존 CELP 코더의 rate-distortion (R-D) 성능을 개선하기 위해서 KLT-CVQ 코더가 제안되었다. 하지만, KLT-CVQ에서는 KLT 변환과 코드북 class 선택을 위해서 EVD를 사용하므로 프레임의 길이가 늘어날수록 계산량이 매우 크게 증가하는 단점이 있었다. 본 논문에서는 EVD 대신에 UTVD를 사용함으로써 훨씬 적은 계산량으로 유사한 R-D 성능을 얻을 수 있는 방법을 제안하였다. 향후에는 계산량 뿐만 아니라 메모리 요구량을 줄일 수 있는 방법에 대한 연구도 진행하고자 한다.

5. 참고문헌

1. M. Y. Kim and W. B. Kleijn, "KLT-based adaptive classified VQ of the speech signal," IEEE Trans. Speech Audio Process., vol.12, no.3, pp.277-289, 2004.
2. G. W. Stewart. "An updating algorithm for subspace tracking." IEEE Trans. Signal Process., vol.40, no.6, pp. 1535-1541, 1992.