음악 음원 분리를 위한 사전정보의 효율적 압축 방식

김민제, 강경옥
한국전자통신연구원 실감음향연구팀

Effective Representation of Prior Knowledge for Music Source Separation

Minje Kim and Kyeongok Kang
Electronics and Telecommunications Research Institute (ETRI), {mkim, kokang}@etri.re.kr

요약

비음성 행렬의 부분적 공동 분해 (Nonnegative Matrix Partial Co-Factorization: NMPCF) 알고리즘을 이용해서 음악 신호로부터 타악기 음원을 분리할 때, 분리 대상 음원의 사전 정보 신호가 대체로 높은 시간 및 공간 복잡도를 야기하는 문제가 있다. 본 논문에서는 이를 해결하기 위한 사전 정보 신호의 효과적 압축 방식을 제안하고, 상용 음악 신호에 대해 적용함으로써 그 성능을 검증한다.

1. 서론

음악 음원 분리는 분리 대상이 되는 음원의 종류에 따라 다양한 응용 가능성을 가지고 있다. 먼저 객체기반 오디오 서비스 및 관련 표준[1]에서는 분리되어 있는 각종 음원을 별도로 제어할 수 있는 사용자 경험을 제공하고 있다. 또한, 보컬 음원 분리 기술은 보다 고품질의 노래방 반주 신호를 효과적으로 제작하는 것에 활용할 수 있으며, 일정 수준의 타악기 분리 기술은 보컬 음원 분리 기술의 성능을 향상시키는 데에 사용될 수 있다[2]. 추가적으로, 타악기 음원과 베이스 기타의 연주 특징은 음악 분위기 인식[3] 등의 음악 분류 문제에서 성능 향상에 기여하고 있다고 보고되었다.

비음성 행렬 분해(Nonnegative Matrix Factorization:NMF)는 혼합 신호에 포함되어 있는 악기 개별의 주파수 성분 및 해당 성분의 시간적 활성화 정도를 분해하는 것에 쓰여 왔다. 최근 NMF의 한 확장 형태인 비음성 행렬의 부분적 공동 분해(Nonnegative Matrix Partial Co-Factorization:NMPCF) 방식은 NMF 기반 음원 분리 방식을 개선하였다[4][5].

본 논문에서는 NMPCF 방식의 음원 분리에서 사전정보를 효과적으로 압축하여 분리 성능을 해치지 않으면서도 연산량을 줄일 수 있는 방식을 제안한다.

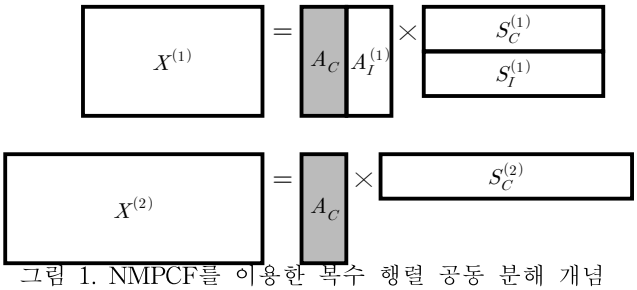

그림 1. NMPCF를 이용한 복수 행렬 공동 분해 개념

2. NMPCF를 이용한 타악기 음원 분리

NMPCF를 이용하는 타악기 음원 분리 방식[4]은 다음과 같은 목적 함수를 최적화함으로써 이루어진다.

$$\mathfrak{I}_{NMPCF} = \sum_{l=1}^{L} \lambda_l \| X^{(l)} - A_C S_C^{(l)} - A_I^{(l)} S_I^{(l)} \|_F^2$$

이를 최적화하는 각 행렬 별 곱셈 갱신 규칙은 다음과 같이 이루어진다.

$$S^{(l)} \leftarrow S^{(l)} \odot \left(\frac{A^{(l)^T} X^{(l)}}{A^{(l)^T} A^{(l)} S^{(l)}} \right),$$

$$A_C \leftarrow A_C \odot \left(\frac{\sum_{l=1}^{L} \lambda_l X^{(l)} S_C^{(l)^T}}{\sum_{l=1}^{L} \lambda_l A^{(l)} S^{(l)} S_C^{(l)^T} + \gamma L A_C} \right),$$

$$A_I^{(l)} \leftarrow A_I^{(l)} \odot \left(\frac{\lambda_l X^{(l)} S_I^{(l)^T}}{\lambda_l A^{(l)} S^{(l)} S_I^{(l)^T} + \gamma A_I^{(l)}} \right)$$

상기 최적화 함수 및 갱신 규칙은 그림 1을 통해 설명될 수 있다. 다양한 타악기의 단독 연주로 이루어진 절대값 스펙트로그램 $X^{(2)}$는 공동 주파수 행렬 A_C의 활성화로 표현된다. NMPCF와 NMF의 다른 점은 A_C가 갱신 과정에서 두 개의 신호 복원에 동시 활용됨으로써, 혼합

신호 $X^{(1)}$의 타악기 음원 역시 복원하는 점이다. 혼합 신호 내 간섭 음원은 개별 주파수 행렬 $A_I^{(1)}$을 통해 표현됨으로써, 공동 주파수 행렬 A_C는 자연히 분류된다. λ_l은 사전 정보 및 혼합 신호 사이의 가중치를 조절한다.

3. 사전정보 신호의 압축

상기 음원 분리 방식은 학습용 사전 정보 행렬 $X^{(2)}$의 크기가 크기 때문에, 갱신 연산에서 많은 계산량을 요구한다. 본 논문에서는 사전 정보 신호 $X^{(2)}$를 짧은 세그먼트 단위로 분할한 다음 $X^{(2)}=[X_1^{(2)},X_2^{(2)},...,X_M^{(2)}]$, 다운믹스하여 축소된 입력 $\widehat{X^{(2)}}=X_1^{(2)}+X_2^{(2)}+...+X_M^{(2)}$를 확보한다. 압축된 사전정보 신호 $\widehat{X^{(2)}}$은 일반적인 오디오 신호 압축과는 달리, 각 세그먼트의 신호가 중첩되어 혼재된 행렬이다. 그러나 본 압축이 유효한 이유는, NMPCF의 덧셈 복원 방식으로 인해, 기존의 방대한 사전 정보 행렬의 분해에 필적한 분해가 가능하기 때문이다.

4. 실험 결과

본 절에서는 상업 음악 신호에서 타악기 음원을 분리한 결과를 제시한다. 분리 성능의 측정을 위해 혼합 직전의 타악기 음원을 확보하였으며, 다음과 같은 신호대 왜곡비(Signal to Distortion Ratio:SDR)를 정의하였다.

$$SDR=10\log_{10}\frac{\Sigma\, s(t)^2}{\Sigma\left(s(t)-\tilde{s}(t)\right)^2}$$

사전 정보 신호와 혼합 신호는 모두 44.1kHz, 16bit PCM 신호이며, 2048 샘플의 프레임을 단구간 퓨리에 변환을 통해 주파수 영역으로 변환하였고, 그 중 7/8 샘플이 중첩되어서 시간 영역의 해상도를 확보하였다. 테스트 신호로는 10곡이 활용되었으며, 곡별로 100초 구간이 10초 단위로 분할되어서 테스트되었다. 사전정보 신호는 테스트에 쓰이지 않는 13 곡에서 선정된 도합 130초 길이의 타악기 연주이다. 실험은 5회 반복 수행되어서 평균치를 비교한다. 압축을 위해서 130초의 사전정보 신호는 10초 길이의 세그먼트로 나누어진 다음 다운믹스되어, 원래 크기의 1/13로 행렬 크기가 줄어들었다.

표 1은 기존의 NMPCF 방식과 압축 방식이 적용된 방식의 결과를 비교하였다. 압축이 적용되면 기존의 사전정보 행렬 분해에 비해 분해 성능이 약간 떨어지지만, 이는 사전정보량을 1/13으로 줄임으로 인해 얻을 수 있는 시간 및 공간 복잡도 측면의 이익에 비해 무시할 수 있는 수준이라고 할 수 있다.

노래	기존 NMPCF 이용 타악기 음원 분리	사전 정보 압축 적용
1	1.9334	1.8844
2	2.7640	2.7754
3	5.1606	5.2768
4	1.9566	1.8936
5	3.5017	3.0817
6	6.3084	6.5145
7	2.7461	2.7621
8	4.2541	4.2257
9	3.2065	3.1473
10	5.3346	5.1911
평균	**3.7166**	**3.6753**

표 1. NMPCF 분리 방식 대비 압축 방식의 분리 성능

5. 결론

본 논문에서는 사전정보 신호를 짧은 세그먼트로 나누어 합치는 압축 방식을 통해 NMPCF를 이용한 음원 분리 성능을 크게 훼손하지 않으면서 시간 및 계산 복잡도를 획기적으로 줄이는 음원 분리 방법이 제시되었다.

감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2010년도 문화콘텐츠산업기술지원사업의 연구결과로 수행되었음.

참고문헌

1. Information technology - Multimedia application format (MPEG-A) - Part12: Interactive music application format, ISO/IEC IS 23 000-12, 2010.

2. M. Kim, I. Jang, and K. Kang, “A unified approach to vocal source separation,” ETRI Journal, (submitted for publication).

3. E. Tsunoo, et al, “Music mood classification by rhythm and bass-line unit pattern analysis,” ICASSP 2010.

4. J. Yoo, M. Kim, K. Kang, and S. Choi, “Nonnegative matrix partial cofactorization for drum source separation,” ICASSP 2010.

5. M. Kim, J. Yoo, K. Kang, and S. Choi, “Blind rhythmic source separation: Nonnegativity and repeatability,” ICASSP 2010.