

위너필터 후처리를 통한 비음수행렬분해 기법의 배경음 저감 성능 향상

이상협* · 김현태**

Improvement of Background Sound Reduction Performance by Non-negative matrix Factorization Method by Wiener Filter Post-processing

Sang Hyeop Lee* · Hyun Tae Kim**

요약

본 논문에서는 비음수 행렬 분해 필터 뒷단에 위너필터를 추가하여 배경음 분리 성능을 향상하는 방법을 제안한다. 배경음이 혼재된 음성 신호의 경우 비음수 행렬 분해 기법으로 1차 분리된 신호에는 아직 완전히 분리되지 못한 부분이 잔류할 수 있다. 이러한 경우 위너필터에 의해 잔류하는 신호의 크기에 비례하여 줄여 줄 수 있어 배경음 분리 또는 저감 효과를 기대할 수 있다. 실험을 통해 위너필터를 추가한 경우가 비음수 행렬 분해 기법만 적용한 경우에 비해 저감 효과가 높은 것을 확인할 수 있었다.

ABSTRACT

In this paper, we propose a method to improve the background sound separation performance by adding a Wiener filter to the end of the non - negative matrix factorization method. In the case of a mixed voice signal with background sound, a part that has not yet been completely separated may remain in the signal that separated first by the non-negative matrix factorization method. In this case, it can be reduced in proportion to the size of the residual signal due to the Wiener filter, so that the background sound separation or reduction effect can be expected. Experimental results show that the addition of the Wiener filter is more effective than the case of applying the non-negative matrix factorization method.

키워드

Nonnegative Matrix Factorization, Wiener Filter, Post Processing, Background Sound Separation
비음수 행렬 분해, 위너 필터, 후처리, 배경음 분리

1. 서론

인터넷의 발달과 다양한 플랫폼들의 등장으로 이제 자국에서 생산되는 영상 콘텐츠 이외에도 다양한 영

상 콘텐츠를 접할 수 있는 기회가 늘었다. 하지만 그 럽에도 불구하고 영상 콘텐츠의 언어는 대부분이 생 산된 나라의 언어로 작성된 경우가 대부분이다. 이 때 번역을 위해 전문 번역사를 이용하면 콘텐츠

* 경성대학교 대학원 전자공학과(keyek23@naver.com) • Received : Jul. 02, 2019, Revised : Jul. 24, 2019, Accepted : Aug. 15, 2019

** 교신저자 : 동의대학교 응용소프트웨어공학전공 • Corresponding Author : Hyun-Tae Kim

• 접수일 : 2019. 07. 02

• 수정완료일 : 2019. 07. 24

• 게재확정일 : 2019. 08. 15

Major of Applied Software Eng., Dongeui University,
Email : htaekim@deu.ac.kr

제작비용이 증가하는 문제가 발생한다. 이에 대한 대안으로 구글 번역기, 네이버 파파고 등 공개 번역기가 많이 활용되고 있다.

공개 번역기는 크게 텍스트 번역과 음성 번역으로 나누어지게 된다. 영상 콘텐츠 특징 상 음성 번역으로 번역이 되어야 하는데 대부분의 영상 콘텐츠에는 콘텐츠의 현장감, 몰입감 등을 고조시키기 위해 배경음이나 효과음이 기본적으로 삽입된다. 이와 같이 배경음이나 효과음이 혼재되어 있는 경우 대사에 대한 음성 인식률은 현저히 떨어진다.

본 논문에서는 이러한 문제를 해결하기 위해 배경음 등이 혼재되어있는 음성에서 배경음을 분리 또는 저감하는 방법을 실시간으로 처리하는 방법을 제안한다. 제안하는 방법은 최근 영상 인식 등의 분야에서 효과가 있다고 알려지고 있는 비음수 행렬 분해 기법(Nonnegative Matrix Factorization)을 활용하여 음성 신호에 혼재된 배경음을 줄이는 방법을 제시한다 [1-4]. 비음수 행렬 분해 기법은 음성 신호의 특징추출과 분류 및 분리를 따로 하는 것이 아닌 스펙트로그램 값을 입력으로 비음수 행렬 분해 기법을 통해 특징을 추출하고 추출된 특징을 이용하여 필터를 설계하여 분류 및 분리하는 방법이다. 그러나 충분히 훈련되지 못한 경우나 특징이 복잡한 경우 분리 효과가 떨어진다. 본 논문에서는 이러한 부분을 개선하기 위해 비음수 행렬 분해 필터 뒷단에 위너필터를 추가하여 배경음 분리 성능을 향상시킨다. 배경음이 혼재된 음성 신호의 경우 비음수 행렬 분해 기법으로 1차 분리된 신호에는 아직 완전히 분리되지 못한 부분이 잔류할 수 있다. 이러한 경우 위너필터에 의해 잔류하는 신호의 크기에 비례하여 잔류 신호를 줄여줄 수 있어 배경음 분리 또는 저감 효과를 기대할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서 기존의 배경음 분리 또는 저감하기 위한 방법에 대해 소개한다. 3장에서는 본 연구의 전반적인 시스템 구성을 알아본 후 비음수 행렬 분해 기법을 이용한 특징추출 및 저감방법에 대해 설명한다. 4장에서는 기존 비음수 행렬 분해 방법과 제안하는 방법과의 성능을 비교하고, 5장에서는 결론을 맺는다.

II. 배경음 분리 또는 저감하는 기존 방법

배경음 분리 또는 저감하는 방법은 기존에도 많이 연구되어 왔던 분야이다. 그러나 기존의 방법들은 특징 추출 단계, 분류 또는 분리 단계 등으로 나누어진 경우가 대부분이다. 따라서 본 논문에서는 기존에 널리 사용되어왔던 특징 추출 알고리즘, 분류 알고리즘과 분리 알고리즘에 대해서 각각 간략히 살펴본다.

2.1 특징 추출 알고리즘

음성 신호의 특징을 추출하는 대표적인 알고리즘은 먼저, 계산량이 적고 비교적 간단한 캡스트럼 방법이 있으며, 음성 특징 추출 분야에서 가장 널리 쓰이는 Mel Frequency Cepstral Coefficient(: MFCC) 방법이 있다. 그 외에도 컨볼루션 신경망을 이용하여 특징을 추출하는 방법도 소개되고 있다.

2.1.1 캡스트럼(Cepstrum)

캡스트럼은 스펙트럼의 이산 푸리에 변환 또는 역변환으로 정의가 된다. 캡스트럼의 용어는 스펙트럼(Spectrum)의 일부 글자를 바꾸어 만들어진 것이며, 캡스트럼의 그래프의 가로축, 즉 종속변수의 단위 또한 frequency라는 단어를 변형시킨 quefrequency라는 단위를 사용한다. 캡스트럼은 주파수 영역의 함수를 변환 또는 역변환 한 것이기 때문에 시간 영역 함수이다. 캡스트럼이 가지는 가장 큰 특징은 음성 신호의 크기와 위상을 분리할 수 있다는 점이다.

2.2.2 Mel Frequency Cepstral Coefficient (MFCC)

MFCC는 사람의 달팽이관이 주파수를 인지하는 방법과 유사한 방식으로써 저주파 대역 에너지보다 고주파 대역 에너지의 차이에 둔감한 원리를 이용한 방식이다[1-2].

2.2 분류 알고리즘

분류 알고리즘에는 신경세포를 모델링한 신경망 알고리즘, 신경망 알고리즘의 문제점을 해결한 다층 신경망 알고리즘, 지도학습에서 사용되는 방법으로 주어진 자료에 대해서 분리하는 초평면을 찾는 Support Vector machines(: SVM) 알고리즘 등이 있다.

2.2.1 신경망

신경망이란 신경세포를 모델링한 뉴런(neuron)이라는 매우 단순한 기능을 가진 유닛들의 상호 연결로 구성되어 일련의 논리 연산이 가능하도록 하는 분류 방법으로써, 연결 강도 및 뉴런의 활성화 함수의 변화를 통해 복수의 선형 또는 비선형 분류기의 결합 형태로 패턴 분류가 이루어진다.

2.2.2 다층 신경망 및 컨볼루션 신경망

다층 퍼셉트론의 경우 XOR 학습이 불가능하고 복잡도가 높은 패턴 분류 분야에서 이용되지 못하는 단점이 있다. 이를 위하여 다층 퍼셉트론이 제안되었으며, 선형 분리만 가능한 다층 퍼셉트론 모델의 단점을 극복하였다. 층의 개수가 증가할수록 퍼셉트론이 구성할 수 있는 결정 구역이 세밀해지며, 일반적으로 델타 규칙 외에 역전파(Backpropagation) 규칙으로 학습한다. 역전파 규칙은 전향 단계와 후향 단계를 반복하여 총 오차의 합이 정해진 오차의 기준치에 도달할 때까지 실행되는 방법으로, 연결 강도의 조절은 일반화한 델타규칙을 적용한다. 전향 단계란 입력 층에 패턴을 입력한 후 활성화 함수를 이용한 출력을 산출하는 단계이며, 후향 단계란 목표 출력과 실제 출력의 오차를 계산한 후 역방향, 즉, 출력 층에서 입력 층으로 층과 층 사이의 연결 강도를 갱신해 나감으로써 오차를 최소화하는 가중치 벡터를 찾아내는 기법이다[5-6].

최근에는 컨볼루션 신경망(CNN)이 매우 활발히 연구되고 있으며, 전통적인 신경망 앞에 여러 계층의 컨볼루션 계층을 붙인 모양인데, 이는 컨볼루션 계층을 통해서 입력 받은 이미지에 대한 특징(Feature)을 추출하게 되고, 이렇게 추출된 특징을 기반으로 기존의 뉴럴 네트워크를 이용하여 분류를 해내게 된다[7-8].

2.2.3 Support Vector Machine(SVM)

SVM은 일반적인 통계적 학습 방법에서의 경험적 리스크 최소화(Empirical Risk Minimization, ERM)와는 달리 구조적 리스크 최소화(Structural Risk Minimization, SRM)을 통해 오류를 최소화 하는 방법으로써, 패턴을 고차원의 특징 공간으로 사상시키는 방법이다[9-10].

2.3 잡음 분리 또는 저감 알고리즘

음성 신호에 섞인 잡음을 분리 또는 저감시키는 알고리즘은 대부분 주파수 영역에서 처리하고 있으며, 대표적인 방법으로 스펙트럼 차감법과 위너필터 방법이 있다.

2.3.1 스펙트럼 차감법

일반적인 주변 잡음은 급작스런 변화없이 존재한다. 따라서 음성이 없는 구간에서도 주변 잡음은 검출할 수 있다. 이러한 특징을 이용하여 주파수 영역에서 음성이 존재하지 않는 구간에서 잡음 스펙트럼을 추정하고 음성이 존재하는 구간에서의 스펙트럼에서 잡음 스펙트럼을 차감하면 잡음이 분리 또는 저감된 음성을 얻을 수 있다. 이러한 방법이 스펙트럼 차감법이다.

2.3.2 위너필터(Wiener filter)

위너 필터는 필터의 출력이 목표 신호와의 오차 신호의 제곱평균 값을 최소화하도록 필터의 계수를 조정해가는 적응필터의 한 예이다. 보통 잡음에 오염된 음성이나 잡음에 오염된 영상신호에 많이 적용되고 있으며, 1차원 신호에 대해 간략히 기술한다. 잡음에 오염된 신호를 $C_i = S_i + N_i$ 라 두면, 오염되지 않은 원래의 음성 신호 추정치를 \hat{S}_i 라 두면, $\hat{S}_i = C_i \phi_i$, $\langle |\hat{S} - S|^2 \rangle$ 즉, 오차 신호의 제곱평균 값을 최소화 하는 필터 ϕ_i 를 찾는 것이 위너 필터를 구하는 과정이다.

$$\begin{aligned} \langle (\hat{S} - S)(\hat{S} - S) \rangle &= \left\langle \sum_i [(S_i + N_i)\phi_i - S_i]^2 \right\rangle \\ &= \sum_i \{ \langle S_i^2 \rangle (1 - \phi_i)^2 + \langle N_i^2 \rangle \phi_i^2 \} - 2 \sum_i \phi_i \langle N_i S_i \rangle \end{aligned} \quad (1)$$

위 식을 ϕ_i 에 대해 미분이 zero 가 되도록 두면, 위너필터는 식 (2)로 구해진다.

$$\phi_i = \frac{\langle S_i^2 \rangle}{\langle S_i^2 \rangle + \langle N_i^2 \rangle} \quad (2)$$

III. 제안하는 방법

3.1 비음수 행렬 분해 방법

비음수 행렬 분해 기법은 오디오 신호의 희소 표현 (Sparse Representation)에 의거하여, 해당 음악 신호를 의미를 가진 여러 개의 음표 성분으로 분리하는 방법으로 주로 사용된다. 또한 특징 추출과 분류기가 따로 있는 것이 아닌 하나의 알고리즘으로 배경음 분리가 가능하다. 비음수 행렬 분해 기법은 다음과 같이 크기의 비음수 행렬 V 를 $K \times R$ 크기의 비음수 행렬 W 와 $R \times N$ 크기의 행렬 H 의 곱으로 나타낸 후, 각 행렬을 추정하여 분해하는 기법이다.

$$V = WH + E \quad (3)$$

일반적으로 오디오 신호 처리 시스템에서 행렬 V 는 음성 신호의 크기 스펙트로그램 (Magnitude Spectrogram)을 사용하며, 이 경우 행렬 W 는 주파수 영역 기저를, 행렬 H 는 시간 영역 기저를 나타낸다.

$$V \in R_+^{F \times T}, \quad W \in R_+^{F \times K}, \quad H \in R_+^{K \times T}, \quad K < F < T \quad (4)$$

이 식들을 이용하여 식 (5)와 같이 유추해낼 수 있다.

$$\begin{bmatrix} | & | & \dots & | \\ V_1 & V_2 & \dots & V_T \\ | & | & \dots & | \end{bmatrix} \approx \begin{bmatrix} \sum_{j=1}^K H_{j1} W_j & \sum_{j=1}^K H_{j2} W_j & \dots & \sum_{j=1}^K H_{jT} W_j \end{bmatrix}$$

$$\begin{bmatrix} | & | & \dots & | \\ V_1 & V_2 & \dots & V_T \\ | & | & \dots & | \end{bmatrix} \approx \begin{bmatrix} | & | & \dots & | \\ W_1 & W_2 & \dots & W_K \\ | & | & \dots & | \end{bmatrix} \begin{bmatrix} -h_1^T - \\ -h_2^T - \\ \vdots \\ -h_K^T - \end{bmatrix}$$

$$V \approx W_1 h_1^T + W_2 h_2^T + \dots + W_K h_K^T \quad (5)$$

위와 같은 비음수 행렬 분해 모델에서 각 기저 행렬, 즉 W 와 H 를 추정하는 기법에는 여러 방법이 있으나, 주로 곱셈 갱신 (multiplicative update) 방법의 의해 순차적으로 추정한다[3-4].

3.2 실시간 비음수 행렬 분해 기법

기존의 비음수 행렬 분해 기법은 모든 데이터를 축적한 뒤 일괄적으로 계산해야 한다는 단점이 있으며, 이를 극복하기 위하여 실시간으로 기저 행렬을 추정하는 알고리즘을 활용한다[3].

실시간 비음수 행렬 분해 기법은 매 n 번째 시간 윈도우마다 다음의 연산을 통해 주파수 기저 행렬 및 시간영역 기저 벡터를 추정한다. 먼저 n 번째 시간 윈도우의 시간영역 기저벡터는 식 (6)과 같이 계산된다.

$$h(n) = [W^\dagger(n-1)v(n)]_+ \quad (6)$$

여기서 $h(n)$ 은 $R \times 1$ 크기의 시간영역 기저 벡터이며, $v(n)$ 은 $K \times 1$ 크기를 가지는 n 번째 시간 윈도우의 크기 스펙트럼 데이터를 나타낸다.

또한 $[\]_+$ 는 0보다 작은 값을 0이 되도록 연산하는 반과 정류 연산을 의미하며, \dagger 는 다음과 같은 무어-펜로즈 의사역행렬 (Moore-Penrose pseudo inverse matrix)를 나타낸다.

$$W^\dagger(n-1) = [W^H(n-1)W(n-1)]^{-1}W^H(n-1) \quad (7)$$

위와 같이 추정된 시간영역 기저 벡터의 값을 이용하여, 주파수 기저 벡터는 다음과 같이 추정된다[8, 9]. 먼저 주파수 빈에 관계없이 공통으로 계산되는 이득 벡터 $k(n)$ 과 시간 기저 벡터 $h(n)$ 의 자기-상관 행렬 (auto-correlation matrix)의 역행렬인 $P(n)$ 은 식 (8)과 같이 계산된다.

$$k(n) = \frac{P(n-1)h(n)}{\lambda + h^T(n)P(n-1)h(n)} \quad (8)$$

$$P(n) = \lambda^{-1}P(n-1) - \lambda^{-1}k(n)h^T(n)P(n-1) \quad (9)$$

여기서 식 (8), (9)의 연산 결과를 이용하여 매 k 번째 주파수 빈마다 식 (10), (11)과 같이 연산을 통해 주파수 기저를 추정한다.

$$\xi_k(n) = v_k(n) - w_k(n-1)h(n) \quad (10)$$

$$w_k^T(n) = [w_k^T(n-1) + k(n)\xi_k^T(n)]_+ \quad (11)$$

여기서 $v_k(n)$ 는 $v(n)$ 의 k 번째 주파수 성분 값을 나타내고, $w_k(n)$ 은 $W(n)$ 행렬의 k 번째 행 벡터를 나타낸다. 그리고 망각 인자 λ 는 일반적으로 1에 가까운 값을 가지며, 과거 값에 대한 가중치를 나타낸다.

3.3 위너필터(Wiener filter) 후처리 연동

비음수 행렬 분해 방법으로 음성과 배경음을 1차적으로 분리되면 아직 완전히 분리되지 못한 잔류 성분이 남아있을 수 있다. 특정 시간 프레임에서 음성 성분이 큰 부분은 큰 만큼 배경음 영역에 남아있을 수 있으며, 같은 이유로 배경음 성분이 큰 시간 프레임에서는 같은 시간 프레임에서의 음성 성분에 잔류할 수 있다. 이러한 사실에 근거하여 비음수 행렬 분해 후 역퓨리어변환 후의 시간 영역에서 2장에 언급한 위너필터를 적용시켜 잔류 성분을 더욱 억제시킨다. 그림 1은 위너필터 후처리를 적용한 전체 시스템의 블록도이다[11].

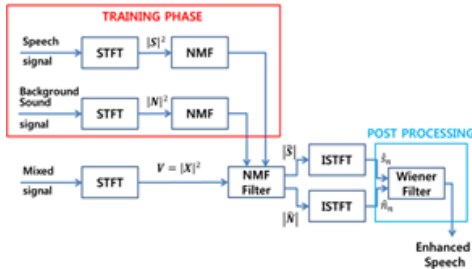


그림 1. 제안하는 방법의 블록도
Fig. 1. Block diagram for the proposed method

IV. 실험 환경 및 결과

4.1 실험 환경

4.1.1 성능 평가 척도

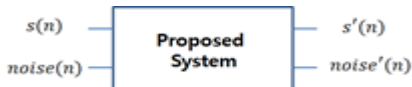


그림 2. SNR 테스트를 위한 블록도
Fig. 2. Block diagram for SNR test

배경음 저감 성능 비교를 위해 다음과 같은 절차에 따라 평가한다. 특정 음성 신호, $s(n)$ 와 특정 배경음(잡음), $noise(n)$ 을 일정비율(SNR 3dB, 혹은 6dB) 섞어 인위적인 혼재 신호 $s_{noise}(n) = s(n) + noise(n)$ 로 만든다. 본 실험에서는 이 배경음(잡음)의 진폭 값을 조정하여 입력 SNR (dB)이 6dB 되도록 하였다.

$$\text{입력 SNR(dB)} = 10 \log_{10} \left(\frac{\sum_N^1 s(n)^2}{\sum_N^1 noise(n)^2} \right) \quad (12)$$

$$\text{출력 SNR(dB)} = 10 \log_{10} \left(\frac{\sum_N^1 s'(n)^2}{\sum_N^1 noise'(n)^2} \right) \quad (13)$$

그 후, 제안하는 배경음 저감 시스템(그림 2 참조)으로 처리 후 출력된 신호 $y(n)$ 에는 순수 음성 성분 $s(n)$ 의 변형 성분인 $s'(n)$ 과 순수 배경음(잡음) 성분 $noise(n)$ 의 변형 성분(또는 잔류 성분)인 $noise'(n)$ 으로 구성되어 있다. 즉, $y(n) = s'(n) + noise'(n)$ 으로 둘 수 있다. 따라서 시스템 출력 SNR은 입력 SNR 구하는 식과 유사한 식(13)와 같이 구한다.

4.1.2 실험에 사용된 데이터

평가를 위한 실험 데이터는 모두 3가지 유형이며 장르 1은 광고 영상의 배경음, 장르 2는 유아 동영상의 배경음, 장르 3는 드럼과 베이스 기타가 섞인 배경음으로 되어 있으며 각 장르별 A, B, C는 5초짜리 음성이다.

4.2 실험 결과

4.2.1 기존 방법과 실시간 비음수행렬분해 방법 비교

먼저 기존 방법과의 비교에서 기존 방법 중 가장 대표적인 방법인 MFCC를 특징 파라미터로 채택하고 분류기로 서포트 벡터 머신(SVM)을 활용한 방법과 비교하였다. 성능 비교 결과는 표 1에 나타내었다. 표 1에서, 비음수 행렬 분해 방법은 기존의 제안되었던

MFCC+SVM 보다 장르에 따라 또한 시간 구간에 따라 적게는 3.234 dB에서 많게는 28.901 dB 차이가 난다. 실험 결과에 대한 원인을 분석해보면, 먼저 MFCC+SVM의 경우 음성과 배경음들 각각의 고차원 특징 값들이 상대적으로 복잡하여 몇 개의 지지벡터 (support vector) 만으로 결정 경계를 나누는 데 한계가 있어 성능이 떨어지는 것으로 판단된다. 또한 배경음 속에 포함된 음원이 복잡할수록 배경음 분리 성능이 떨어지는 것도 확인할 수 있었다. 장르 1의 경우는 음원 구성이 상대적으로 유아 동영상의 배경음이나 드럼과 베이스 기타가 섞인 배경음 보다 다소 복잡한 것으로 분석되었다.

표 1. 기존 대표 방법과 실시간 NMF간 성능 평가 결과

Table 1. Performance evaluation results between the proposed method and conventional NMF

Method File name	MFCC+SV M(dB)	NMF(dB)
Genre 1 - A	11.651	16.251
Genre 1 - B	15.162	18.623
Genre 1 - C	17.929	26.326
Genre 2 - A	2.636	22.746
Genre 2 - B	6.753	19.152
Genre 2 - C	1.162	20.191
Genre 3 - A	32.362	61.263
Genre 3 - B	6.238	11.436
Genre 3 - C	10.002	13.236

4.2.2 기존 비음수 행렬 분해 방법과 위너필터 후처리 연동 비음수 행렬 분해 방법 비교

두 번째는 기존 비음수 행렬분해 방법과 제안하는 위너필터 후처리를 추가한 방법과의 비교이다. 성능 비교 결과는 표 2에 나타내었다. 표 2에서 장르와 시간 구간에 따라 적게는 0.005 dB에서 많게는 4.727 dB 차이가 난다. 이것은 3.3절에서도 언급한 바와 같이 1차적으로 비음수 행렬 분해 방법으로 분리되더라도 완전히 분리되지 못한 잔류 성분이 존재하고 위너필터를 통해 잔류 성분이 더욱 줄어들었다는 것을 나타내고 있다고 판단된다.

표 2. 후처리 연동 방법과 기존 NMF간 성능 평가 결과
Table 2. Performance evaluation results between the proposed method and conventional NMF

Method File name	NMF(dB)	NMF+Wiener filter(dB)
Genre 1 - A	16.251	17.460
Genre 1 - B	18.623	19.162
Genre 1 - C	26.326	26.331
Genre 2 - A	22.746	22.862
Genre 2 - B	19.152	21.126
Genre 2 - C	20.191	20.236
Genre 3 - A	61.263	62.623
Genre 3 - B	11.436	16.163
Genre 3 - C	13.236	13.563

V. 결론

본 논문에서는 음성에 혼재되어있는 효과음이나 배경음을 효과적으로 분리 또는 저감하는 방법으로 비음수 행렬 분해 방법과 연동한 후처리 방법을 제안하고 기존 분리방법 및 기존 비음수분해방법과 비교를 통해 제안하는 방법이 효과가 있다는 것을 제시하였다. 실험을 통해 음원의 구성이 복잡한 배경음의 경우에 분리 또는 저감 성능이 다소 떨어지는 것도 확인할 수 있었다. 향후는 보다 복잡한 음원으로 구성된 배경음의 경우에 대해서도 분리 성능을 높이는 연구를 지속할 예정이다.

감사의 글

본 논문은 2015년도 동의대학교 연구년 지원에 의하여 연구되었음.

References

- [1] H. Lee, J. Seo, Y. Lee, J. Kim, and S. Lee, "Study on optimal number of latent source in speech enhancement based Bayesian nonnegative matrix factorization," In *Proceeding of 2015 The Korean Institute of Broadcast and Media Engineering Conference*, Jeju

island, South Korea, July 2015, pp. 418-420.

[2] J. Yoo, M. Kim, K. Kang, and S. Choi, "Nonnegative Matrix Partial Co-factorization for Drum Source Separation," *2010 IEEE International Conference on ASSP(Acoustics, Speech, and Signal Processing)*, Dallas, Texas, USA, Mar. 2010.

[3] S. Lee, "Online Monaural Ambient Sound Extraction based on Nonnegative Matrix Factorization Method for Audio Contents," *Journal of The Korean Institute of Broadcast and Media Engineering*, vol. 19, no. 6, Nov. 2014, pp. 819-825.

[4] S. Lee, "Research on Reducing Background Sounds in Mixed Speech Signals," Master Thesis, *Donggeui University*, 2018.

[5] R. Parekh, J. Yang, and V. Honavar, "Constructive neural network learning algorithm for pattern classification," *J. of IEEE Transactions on Neural Networks*, vol. 11, no. 2, 2000, pp. 436-451.

[6] D. Liu, T. Chang, and Y. Zhang "A new learning algorithm for feed forward neural networks," In *Proceedings of IEEE International Symposium on Intelligent Control*, Mexico City, Mexico, 2001, pp. 39-44.

[7] G. Son and J. Park, "A Study of Kernel Characteristics of CNN Deep Learning for Effective Fire Detection Based on Video," *J. of the Korea Institute of Electronic Communication Science*, vol. 13, no. 6, 2018, pp. 1257-1262.

[8] J. Park, M. Wiranegara, and G. Son, "Multi-channel Video Analysis Based on Deep Learning for Video Surveillance," *J. of the Korea Institute of Electronic Communication Science*, vol. 13, no. 6, 2018, pp. 1263-1268.

[9] V. Vapnik, "An overview of statistical learning theory," *J. of IEEE Transactions on Neural Networks*, vol.10, issue 5, 1999, pp. 988-999.

[10] M. M. Adankon, M. Cheriet, and A. Biem, "Semisupervised Learning Using Bayesian Interpretation : Application to LS-SVM," *J. of IEEE Transactions on Neural Networks*, vol. 22, issue. 4, 2011, pp. 513-524.

[11] S. Lee, J. Park, and H. Kim, "Reduction of background sound on non-negative matrix factorization combined with Wiener filter

post-processing," *Proceedings of 18th International Conference on Control, Auto-mation and Systems(ICCAS2018)*, Pyeongchang, South Korea, Oct. 2018.

저자 소개



이상협(Sang-Hyeop Lee)

2016년 동의대학교 멀티미디어공학과 졸업(공학사)

2018년 동의대학교 대학원 디지털미디어공학과 졸업(공학석사)

2018년 ~ 현재 경성대학교 대학원 전기전자공학과 박사과정

※ 관심분야 : 영상 및 음향신호처리, 컴퓨터비전, 딥러닝



김현태(Hyun-Tae Kim)

1989년 부산대학교 전자공학과 졸업(공학사)

1995년 부산대학교 대학원 전자공학과 졸업(공학석사)

2000년 부산대학교 대학원 전자공학과 졸업(공학박사)

2002년~현재 동의대학교 응용소프트웨어공학전공 교수

※ 관심분야 : 영상 및 음향신호처리, 컴퓨터비전, 딥러닝

