

# Bag of Words 기반 음향 상황 인지를 위한 주파수-캡스트럴 특징

## Frequency-Cepstral Features for Bag of Words Based Acoustic Context Awareness

박상욱, 최우현, 고한석<sup>†</sup>

(Sang-wook Park, Woo-hyun Choi, and Hanseok Ko<sup>†</sup>)

고려대학교 전기전자전파공학과

(접수일자: 2013년 12월 19일; 수정일자: 2014년 4월 21일; 채택일자: 2014년 6월 2일)

**초 록:** 음향 상황 인지(acoustic context awareness)는 다양하게 발생하는 음원들로부터 어떠한 장소인지 또는 어떠한 사건이 발생하는지를 판단하는 기술로 음향 이벤트 검출 또는 인식 보다 한 단계 더 복잡한 문제이다. 기존의 상황 인지 기술은 음향 이벤트 검출 또는 인식 기술에 기반하여 현재 상황을 인지하는 방법을 사용하고 있다. 하지만 이와 같은 접근 방법은 여러 음원이 동시에 발생하거나 유사한 음원이 발생하는 실제 환경에서 정확한 상황 판단이 어렵다. 특히 버스와 지하철은 승객들에 의한 잡음으로 상황을 인지하기 힘들다. 이러한 문제를 극복하기 위해 본 논문에서는 유사한 음향 이벤트가 발생하는 버스와 지하철 상황을 인식할 수 있는 Bag of Words 기반의 상황 인지 알고리즘을 연구하고 코드북 생성을 위한 특징벡터를 제안한다. 제안하는 특징벡터의 효용성은 Support Vector Machine을 이용한 실험을 통해 검증했다.

**핵심용어:** 음향 상황 인지, 음향 이벤트 검출 또는 인식, Bag of Words (BOW), Support Vector Machine (SVM)

**ABSTRACT:** Among acoustic signal analysis tasks, acoustic context awareness is one of the most formidable tasks in terms of complexity since it requires sophisticated understanding of individual acoustic events. In conventional context awareness methods, individual acoustic event detection or recognition is employed to generate a relevant decision on the impending context. However this approach may produce poorly performing decision results in practical situations due to the possibility of events occurring simultaneously or the acoustically similar events that are difficult to distinguish with each other. Particularly, the babble noise acoustic event occurring at a bus or subway environment may create confusion to context awareness task since babbling is similar in any environment. Therefore in this paper, a frequency-cepstral feature vector is proposed to mitigate the confusion problem during the situation awareness task of binary decisions: bus or metro. By employing the Support Vector Machine (SVM) as the classifier, the proposed feature vector scheme is shown to produce better performance than the conventional scheme.

**Keywords:** Acoustic context awareness, Acoustic event detection or recognition, Bag of Words (BOW), Support Vectors Machine (SVM)

**PACS numbers:** 43.60.Bf

### 1. 서 론

음향 상황 인지(acoustic context awareness)는 다양

한 음원이 발생하는 환경에서 현재 어떠한 장소에 있는지 혹은 어떠한 사건이 발생하였는지를 판단하는 기술이다. 사용자가 항상 소지하고 다니는 스마트폰에 상황 인지 알고리즘을 적용하면 사용자에게 여러 편의를 제공할 수 있다.<sup>[1]</sup>

음향 이벤트 검출 또는 인식에 기반하여 음향 상

<sup>†</sup>Corresponding author: Hanseok Ko (hsko@korea.ac.kr)  
Engineering Building, Room 419, Department of Electronics and  
Computer Engineering, Korea University Anam Campus, Anam-  
Dong 5-Ga, Seongbuk-Gu, Seoul 136-713, Republic of Korea  
(Tel: 82-2-3290-3239)

황을 인지를 수행하는 기존 연구에서 Mel Frequency Cepstral Coefficient(MFCC)와 Hidden Markov Model(HMM)을 사용하였다.<sup>[2,6]</sup> 또한 Linear Predictive Coding Cepstrum(LPCC)와 HMM을 사용하여 5가지 환경음을 인식한 연구도 발표되었다.<sup>[7]</sup> HMM이외에 Support Vector Machines(SVM)을 사용하여 음향 이벤트를 인식한 연구도 발표되었다.<sup>[8,9]</sup> 감시 시스템을 위한 비정상 상황 인지 기술에서는 MFCC와 음색특징을 사용하여, Hierarchical Gaussian Mixture Model(Hierarchical-GMM)로 인식을 수행했다.<sup>[10]</sup> 최근 음원 분리 기술을 적용하여 중첩된 이벤트가 발생하는 상황에서도 인식을 수행하고 있다.<sup>[11,12]</sup> 하지만 기존의 접근 방법으로는 동일한 이벤트가 발생하는 서로 다른 상황을 인지하기 어렵다.

본 논문에서는 동일한 이벤트가 발생하는 버스와 지하철 상황을 인지하기 위해 Bag of Words (BOW) 기반의 인식 알고리즘을 적용하였다. BOW 기반 인식 알고리즘은 코드북 기반의 히스토그램 특징을 이용하여 인식을 수행하기 때문에 동일한 이벤트가 발생하는 상황에서도 상황을 인식할 수 있다. 이때, 코드북의 표현력을 결정하는 코드북 크기와 코드북을 구성하는 특징에 관한 이슈가 있다.<sup>[13]</sup> 코드북이 각 상황의 대표적인 특징을 효과적으로 반영하도록 주파수-캡스트럴 특징을 제안하였고, 코드북 크기에 따라 상황 인지 실험을 수행하였다. 실험을 통해 제안하는 특징을 사용한 경우 동일한 이벤트가 발생하는 상

황을 인지하는데 가장 높은 인식 성능을 확인하였다.

본 논문은 II장에서 BOW 기반 음향 상황 인지의 흐름도(flowchart)와 제안하는 특징을 설명하고, III장에서 기존의 방법들과 제안하는 방법을 이용한 실험에 대해 다루며, IV장에서 결론으로 구성되었다.

## II. BOW기반 음향 상황 인지

이벤트 인식에 기반한 기존의 상황 인지 알고리즘은 잡음과 여러 이벤트가 중첩되어 발생하는 환경에서 이벤트 인식률을 보장하기 위해 잡음 제거 또는 음원 분리 알고리즘이 요구된다. 특히 버스와 지하철 상황은 승객들로 인해 말소리, 기침, 웃음소리, 비닐봉지 소리 등 공통된 이벤트들이 발생하기 때문에 두 상황을 인지하는데 어려움이 있다. 반면, BOW기반 알고리즘은 코드북과 유사도 비교를 통해 가장 유사한 요소로 대체되기 때문에 비교적 왜곡에 강인하다. 또한 유사한 이벤트가 발생하는 환경에서 일정 시간 동안 발생한 특징의 분포를 고려함으로써 문제를 해결할 수 있다.

이러한 BOW기반 상황 인지 알고리즘의 성능은 코드북의 크기와 이를 구성하는 특징에 따라 다르다.<sup>[13]</sup> 코드북의 크기가 크고 이를 구성하는 벡터가 음향적 특징을 효과적으로 반영할수록 다양한 경우를 포괄할 수 있어 높은 성능을 기대할 수 있다.

Fig. 1은 BOW기반 음향 상황 인지 알고리즘의 흐

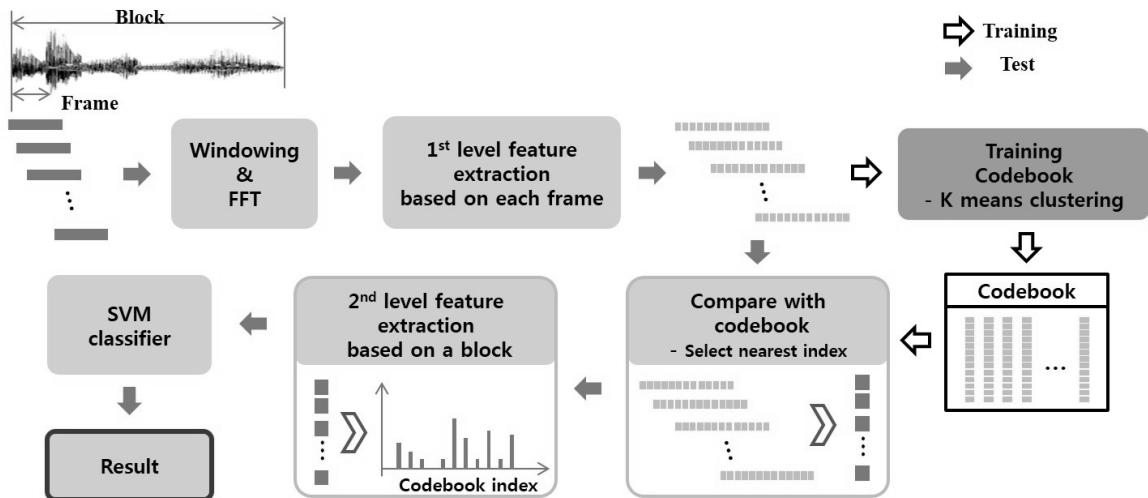


Fig. 1. Flowchart of acoustic context awareness using the Bag of Words approach.

Table 1. Frequency feature set.

Feature	Description	Equation
Spectral Centroid (SC) <sup>[15]</sup>	Center of mass of the spectrum	$SC = \frac{1}{\sum_{i=1}^{L_F}  X(i) } \sum_{i=1}^{L_F} i \frac{F_s}{L_F}  X(i) $
Spectral Spread (SS) <sup>[14]</sup>	Standard deviation of the SC	$SS = \sqrt{\frac{1}{\sum_{i=1}^{L_F}  X(i) } \sum_{i=1}^{L_F} (i \frac{F_s}{L_F} - SC)^2  X(i) }$
Spectral roll-off ( $f_{ro}$ ) <sup>[10]</sup>	Measure of the skewness of the spectrum	$f_{ro} = \min_{f_m} \{f_{ro} \mid \sum_{i=1}^{f_m}  X(i)  \geq 0.9 \sum_{i=1}^{L_F}  X(i) \}$
Energy Ratio in Sub-band (ERSB) <sup>[14]</sup>	Energy distribution rough approximation of a spectrum (4 sub-bands)	$ERSB_n^k = \frac{1}{\sum_{i=1}^{L_F}  X_n(i) ^2} \sum_{i=u_k}^{u_{k+1}}  X_n(i) ^2$ where $\begin{cases} l_1 = 1, u_1 = \lfloor L_F * 0.028 \rfloor \\ l_2 = \lfloor L_F * 0.028 \rfloor + 1, u_2 = \lfloor L_F * 0.078 \rfloor \\ l_3 = \lfloor L_F * 0.078 \rfloor + 1, u_3 = \lfloor L_F * 0.2 \rfloor \\ l_4 = \lfloor L_F * 0.2 \rfloor + 1, u_4 = \lfloor L_F * 1 \rfloor \end{cases}$ $\lfloor x \rfloor : \text{the largest integer not greater than } x$
Zero Crossing Rate (ZCR) <sup>[15]</sup>	The rate of sign-changes along a frame	$ZCR = \frac{1}{2L} \sum_{i=1}^L  sgn(x(i+1)) - sgn(x(i)) $ where $sgn(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ -1, & \text{otherwise} \end{cases}$

$F_s$ : sampling rate,  $L_F$ : maximum FFT bin, L: frame length,

$X(i)$ : i-th magnitude of a frame in frequency domain,  $x(i)$ : i-th amplitude of a frame in time domain.

를 나타낸다. 블록은 인식 결과를 도출하는 주기를 뜻하고, 한 블록 안에는 여러 frame이 존재한다. 프레임마다 1<sup>st</sup> level 특징을 추출하고 사전에 훈련된 코드북과 비교하여 가장 유사한 index를 찾는다. 이 index의 분포로써 2<sup>nd</sup> level 특징을 추출한다. 인식 단계에서는 Support Vector Machine(SVM)을 이용하여 인식 결과를 도출한다.

본 논문에서는 BOW기반 음향 상황 인지에서 코드북을 구성하는 1<sup>st</sup> level 특징 벡터로써 주파수-캡스트럴 특징을 제안한다.

## 2.1 1<sup>st</sup> level 특징 추출: 주파수-캡스트럴 특징

### 2.1.1 주파수 특징 벡터(Frequency Feature Vector)

음색 정보는 주로 음악의 장르를 인식하기 위한 특징으로 사용된다.<sup>[15]</sup> 음색은 주파수에 따른 에너지 분포와 밀접한 관계가 있기 때문에 음향 특징을 반영하기 위해 사용할 수 있다. 최근 비정상 상황인지를 위한 논문 중에도 음색 정보를 사용하여 신뢰할 만한 성능을 보이고 있다.<sup>[10,14]</sup>

Table 1에 논문에서 사용한 음색 특징을 정리했다. 본 논문에서는 frame 에너지 E, frame 에너지 변화량  $\Delta E$ , 에너지의 실효값을 의미하는 volume V지표를

위 음색 특징과 함께 사용했으며, 수식은 각각 다음 Eqs.(1)~(3)과 같다.

$$E = \sum_{i=1}^L x^2(i). \quad (1)$$

$$\Delta E = E_n - E_{n-1}. \quad (2)$$

$$V = \sqrt{\frac{1}{L} \sum_{i=1}^L x^2(i)}. \quad (3)$$

위 Eq.(2)에서  $E_n$ 은 n번째 frame에서 에너지를 뜻한다.

### 2.1.2 캡스트럴 특징 벡터(Cepstral Feature Vector)

캡스트럴 특징을 사용하면 스펙트럼(spectrum)의 포락선(envelope)을 반영할 수 있다. Fig. 2는 논문에서 제안하는 캡스트럴 특징 추출 과정을 나타내며 필터링(filtering) 단계를 제외하고 MFCC를 추출하는 과정과 동일하다. 본 논문에서는 필터링과정에서 버스와 지하철 상황에서 에너지 분포가 유사한 대역에서 높은 분해능을 갖는 필터뱅크(filterbank)를 사용했다. 따라서 에너지 분포가 유사한 대역에서 스펙트럼의 포락선 차이를 효과적으로 반영할 수 있다.

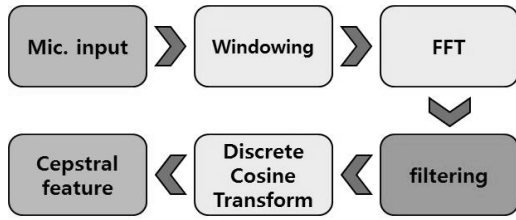


Fig. 2. A procedure for cepstral feature extraction.

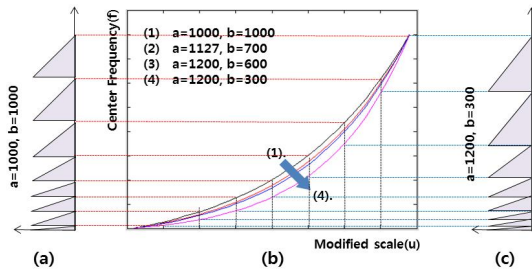


Fig. 3. Filterbank according to constants. (a) Filterbank when a=1000, b=1000, (b) Center frequency according to constant a, b, (c) Filterbank when a=1200, b=300.

필터뱅크의 분해능은 각 필터의 중심 주파수의 위치에 따라 다르다. 본 논문에서는 Eq.(4)를 이용하여 필터뱅크의 중심주파수를 설계했다.

$$f(u) = b \times (\exp(u/a) - 1). \tag{4}$$

Eq.(4)에서 u는 필터뱅크의 index를 뜻하며, f(u)는 u번째 필터의 중심 주파수를 뜻한다. a와 b는 상수 값으로 이를 조절함으로써 f(u)를 조절할 수 있다. Fig 3은 a, b 두 상수에 따라 설계된 필터뱅크를 보여준다.

필터뱅크의 중심주파수가 밀집된 대역에서 스펙트럼의 포락선이 효과적으로 반영됨으로써 높은 분해능을 갖게 된다. 버스와 지하철에서 공통으로 발생하는 이벤트들은 사람 말소리(babble), 휴대전화 벨소리, 안내방송 등이 있다. 이러한 이벤트들은 공통적으로 약 2 kHz 이상의 주파수 대역에서 에너지가 분포하기 때문에 해당 대역의 분해능을 낮추어 유사한 이벤트로 발생하는 오류를 줄일 수 있다. 본 논문에서는 필터뱅크에서 두 변수 a, b에 따라 총 38개의 필터를 사용하였으며, HTK<sup>[11]</sup>를 이용하여 추출한 MFCC와 성능 비교를 위해 Fig. 3과 같이 삼각모양의 필터를 사용하였다.

## 2.2 코드북 생성

코드북 생성에는 훈련데이터로부터 frame 기반의 1<sup>st</sup> level 특징이 사용된다. 훈련데이터는 각 frame에서 주파수-캡스트럴 특징을 추출한 뒤 K-means 군집화를 통해 K개의 군집으로 분할된다. 이때, 각 군집의 평균 벡터들을 이용하여 코드북을 구성한다. K는 코드북의 크기를 결정하는 동시에 2<sup>nd</sup> level 특징의 차원이 된다. K가 클수록 상황을 대표하는 특징 벡터의 수가 증가하여 코드북의 표현력이 향상되지만 인식에 사용되는 특징벡터의 차원은 증가한다.

## 2.3 2<sup>nd</sup> level 특징 추출: 히스토그램 특징

인식에 사용할 특징 벡터로 일정기간(블록) 동안 발생한 1<sup>st</sup> level 특징의 분포를 사용한다. 앞서 추출된 frame 기반의 주파수-캡스트럴 특징은 코드북과 유클리디안 거리를 사용하여 가장 유사한 코드북의 index로 변환된다. 블록 내 모든 frame에 대해서 코드북의 index로 변환한 뒤 index에 대한 히스토그램으로 나타낼 수 있다. BOW를 이용하면 훈련에 사용되지 않은 입력벡터라도 코드북에서 가장 유사한 특징으로 대체되기 때문에 다양한 이벤트가 발생하는 상황도 적은 훈련 데이터를 사용하여 인식할 수 있다.

## 2.4 Support Vector Machine (SVM)

본 논문에서는 특징벡터 차원(코드북 군집 수)이 크고 대부분 0의 요소를 갖기 때문에(블록 내 frame 수 << 코드북 군집 수) SVM을 이용한 방법이 인식에 효율적이다. SVM은 커널 함수에 따라 성능이 달라질 수 있기 때문에, 본 논문에서는 네 가지 커널함수를 사용하여 성능을 비교 측정하였다.

# III. 실험

## 3.1 실험 database

본 논문에서는 유사한 환경음 속에 동일한 음향 이벤트가 발생하는 상황을 인식하기 위해 버스와 지하철 상황을 선정 하였다. Table 2는 실험에 사용된 DB의 분량을 보여준다. 실험 database는 녹음기를 이용하여 서울 시내 일반버스(Normal)와 저상버스(Low),

Table 2. Database description.

Bus	Normal	6 h 35 m 57 s	Low	1 h 19 m 10 s
Metro	Line no.1	1 h 52 m 29 s	Line no.7	1 h 46 m 05 s
	Line no.2	1 h 18 m 22 s	Line no.8	54 m 09 s
	Line no.3	1 h 54 m 09 s	Line no.9	1 h 11 m 12 s
	Line no.4	2 h 03 m 39 s	Bundang line	1 h 12 m 40 s
	Line no.5	2 h 03 m 24 s	Airport line	47 m 28 s
	Line no.6	3 h 30 m 58 s		

그리고 서울 지하철(1호선~9호선과 분당선, 공항철도)에서 수집하였다. 녹음 데이터는 실험을 위해 16 kHz, Mono, 16 bits resolution으로 변환하였다.

코드북 생성과 SVM 훈련을 위한 데이터는 버스와 지하철의 주행 잡음 구간으로 약 2,000s 분량을 구축하였다. 테스트 데이터는 한 블록 단위(4s)로 분할하여 총 15,448개의 database를 구축하였다. 테스트 데이터는 실제 상황과 동일하게 대화소리, 기침 등 승객들로 인해 발생할 수 있는 이벤트들과 라디오, 안내방송, 버스 경적음, 버스카드부저음, 버스 하차벨음, 버스 또는 지하철 주행음 등 다양한 이벤트들이 포함되어 있다.

### 3.2 Baseline 1: MFCC-SVM 실험

첫 번째 baseline은 BOW를 적용하지 않고 MFCC를

이용한 성능이다. Frame 길이는 32 ms, 50% 중첩하여 MFCC 특징 벡터를 추출하였다. 이때 특징 벡터는 11 차원의 MFCC와 Delta특징으로 총 22차원 벡터를 사용했다. SVM 훈련에서 사용된 커널함수는 선형(linear) 커널, 2차형(quadratic) 커널, Gaussian Radial Basis Function(GRBF), Multilayer Perceptron Kernel(MLP)로 4개의 커널 함수를 적용했다.

### 3.3 Baseline 2: BOW 기반 음향 상황 인지

참고문헌 14 논문에서는 코드북 구성을 위해 SC, SS, roll-off, ERSB, ZCR, E, V의 주파수 특징을 사용하였다. 실험은 코드북 크기에 따라 128, 256, 512 총 세 가지 경우에 대해서 진행했으며, 인식을 위한 SVM은 MFCC-SVM 실험과 동일한 커널 함수에서 성능을 각각 측정했다.

### 3.4 제안하는 방법

BOW를 위해 주파수-캡스트럴 특징을 이용하여 코드북을 구성했다. 캡스트럴 특징은 22차원 벡터를 사용했고, 두 번째 baseline과 동일하게 코드북의 크기에 따라 실험을 진행했다. 두 baseline 실험 결과와 제안하는 방법의 실험 결과를 Table 3에 정리했다.

Table 3. Experimental results for acoustic context awareness.

Experiment description	codebook size	parameter	SVM kernel function			
			Linear	Quadratic	GRBF	MLP
Baseline 1 (MFCC)	-	-	No converge	68.53 %	45.02 %	<b>75.60 %</b>
Baseline 2 [Frequency(BOW)]	512	-	No converge	<b>83.80 %</b>	55.39 %	71.45 %
	256	-	78.85 %	<b>82.82 %</b>	55.40 %	34.07 %
	128	-	No converge	<b>71.80 %</b>	56.40 %	45.00 %
Proposed [frequency - cepstral (BOW)]	512	a=1127, b=700	<b>91.18 %</b>	90.69 %	55.54 %	45.03 %
		a=1000, b=1000	<b>91.27 %</b>	90.63 %	55.46 %	44.96 %
		a=1200, b=600	<b>89.27 %</b>	88.50 %	76.26 %	55.55 %
		a=1200, b=300	<b>88.63 %</b>	87.18 %	55.42 %	47.59 %
	256	a=1127, b=700	88.89 %	<b>89.82 %</b>	56.01 %	68.60 %
		a=1000, b=1000	88.40 %	<b>88.74 %</b>	55.97 %	52.89 %
		a=1200, b=600	<b>89.40 %</b>	88.91 %	56.00 %	71.65 %
		a=1200, b=300	<b>87.86 %</b>	87.82 %	55.96 %	50.30 %
	128	a=1127, b=700	<b>90.10 %</b>	87.34 %	64.29 %	58.46 %
		a=1000, b=1000	<b>89.34 %</b>	87.05 %	59.68 %	74.79 %
		a=1200, b=600	<b>87.18 %</b>	83.10 %	59.35 %	75.83 %
		a=1200, b=300	<b>85.46 %</b>	85.10 %	61.98 %	80.31 %

### 3.5 실험 결과 분석

본 실험에서는 코드북 크기, 필터뱅크 매개변수 (parameter), SVM 커널함수에 따른 인식 실험을 진행하였다. BOW 기반 음향 상황인지 알고리즘에서 코드북 크기가 클수록 다양한 음향 상황을 표현할 수 있어 높은 성능을 확인할 수 있다. 제안하는 방법을 적용한 실험에서 코드북 크기가 동일하더라도 코드북을 구성하는 특징 벡터에 따라 성능이 달라짐을 확인할 수 있다. 2kHz 보다 낮은 주파수 대역에 높은 분해능을 갖는 필터뱅크를 사용한 경우( $a=1127, b=700$  또는  $a=1000, b=1000$ )에 유사한 성능을 확인할 수 있으며, 2kHz보다 높은 주파수 대역에 높은 분해능을 갖는 경우( $a=1200, b=600$  또는  $a=1200, b=300$ )보다 높은 성능을 보임을 확인할 수 있다.

#### 3.5.1 MFCC-SVM

Table 3에서 MFCC 특징벡터를 사용한 경우, MLP 커널을 사용했을 때 인식률이 가장 높다. 훈련 데이터가 주행 잡음으로만 구성되어 있기 때문에 다양한 이벤트가 복합적으로 발생하는 상황에서 오인식이 많이 발생했다. 만일 훈련 데이터에 복합적인 상황의 데이터가 포함되었다면 인식률이 더 상승할 것이다. 하지만 그 만큼 많은 훈련 데이터가 필요하다.

#### 3.5.2 BOW 기반 음향 상황 인지

이번 경우에는 2차형 커널에서 인식률이 가장 높다. 14에선 이벤트가 중첩된 상황이 아닌 단독으로 발생하므로 주파수 특징만으로 인식 대상을 표현할 수 있지만, 다양한 이벤트가 발생하는 버스 또는 지하철 상황을 표현하기에는 부족하다. 하지만 BOW 접근법을 적용하여 다양한 상황에 대응할 수 있어 기존의 MFCC보다 높은 성능을 보이고 있다.

#### 3.5.3 제안하는 방법

코드북을 구성하는 벡터가 각 음향 상황을 효과적으로 표현하고 있기 때문에 상황인지를 위한 히스토그램 특징이 버스와 지하철에 따라 차별적으로 추출된다. 실험 결과에서도 대부분 1차 선형 커널 함수를 사용한 경우에 가장 높은 인식률을 확인할 수 있다.

Table 3의 결과들을 바탕으로, 주파수-캡스트럴 특

징이 공통으로 발생하는 이벤트의 영향을 줄이고 복합적인 상황을 더욱 자세히 반영할 수 있기 때문에 BOW 기반 음향 상황 인지를 위한 특징벡터로 적합하다고 할 수 있다.

## IV. 결 론

기존 음향 상황 인지를 위한 알고리즘은 이벤트가 중첩되거나 유사한 이벤트가 발생하는 상황에서는 어려움이 있다. 이러한 문제점을 해결하기 위해 본 논문에서는 주파수-캡스트럴 특징 기반의 BOW 접근법을 사용하였다. 본 논문에서 제안하는 특징은 주파수에 따른 에너지 분포와 스펙트럼의 포락선을 반영하고 있으며 동시에 본 논문에서 사용한 캡스트럴 특징은 에너지 분포가 유사한 대역에서 높은 분해능을 갖는 필터뱅크를 사용함으로써 포락선을 효과적으로 반영할 수 있다. SVM을 이용한 비교 실험을 통해 제안한 방법을 적용한 경우 기존 방법들보다 높은 성능을 확인하였다.

## 감사의 글

본 연구는 서울시 산학연 협력사업(WR080951)의 연구결과로 수행되었습니다.

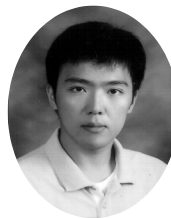
## References

1. W.H. Choi, S.I. Kim, M.S. Keum, D.K. Han, and H. Ko, "Acoustic and visual signal based context awareness system for mobile application," IEEE Trans. Cons. Elec. **57**, 2 738-746 (2011).
2. B. Clarkson, N. Sawhney, and A. Pentland, "Auditory context awareness via wearable computing," in Proc. Works. Perceptual User Interface, 37-42 (1998).
3. L. Ma, B. Milner, and D. Smith, "Environmental noise classification for context-aware application," in Proc. Works. Database and Expert Sys. Appl. **2736**, 360-370 (2003).
4. L. Ma, B. Milner, and D. Smith, "Acoustic environment classification," ACM Trans. Speech and Lang. Process. **3**, 2, 1-22 (2006).
5. A. J. Eronenm, V. T. Peltonen, J. T. Tuomi, A. P. Klapuri, S. Fagerlund, T. Sorsa, G. Lorho, and J. Huopaniemi,

- “Audio-based context recognition,” IEEE Trans. Audio, Speech, and Lang. Process. **14**, 321-329 (2006).
6. T. Nishiura, S. Nakamura, K. Miki, and K. Shikano, “Environmental sound source identification based on hidden Markov model for robust speech recognition,” in Proc. Eurospeech-Interspeech, 2157-2160 (2003).
  7. P. Gaunard, C.G. Mubikangiey, C. Couvneur, and V. Fontaine, “Automatic classification of environmental noise events by hidden Markov model,” in Proc IEEE Inter. Conf. Acoust., Speech, and Sig. Process. **6**, 3609-3612 (1998).
  8. G. Guo and S.Z. Li, “Content-based audio classification and retrieval by support vector machines,” IEEE Trans. Neural Networks **14**, 209-215 (2003).
  9. A. Temko, E. Monte, and C. Nadeu, “Comparison of sequence discriminant support vector machines for acoustic event classification,” in Proc. IEEE Inter. Conf. Acoust., Speech, and Sig. Process. **5**, 721-724 (2006).
  10. K. Kim and H. Ko, “Hierarchical approach for abnormal acoustic event classification in an elevator,” in Proc. IEEE Inter. Conf. Ad. Video and Sig. Surveillance, 88-94 (2011).
  11. T. Heittola, A. Mesaros, A. Eronen, and T. Virtanen, “Context-dependent sound event detection,” EURASIP J. Audio, Speech, and Music Process. **1**, 1-13 (2013).
  12. T. Heittola, A. Mesaros, T. Virtanen, and M. Gabbouj, “Supervised model training for overlapping sound events based on unsupervised source separation,” in Proc IEEE Inter. Conf. Acoust., Speech, and Sig. Process. 8677-8681 (2013).
  13. S. Rawat, P. F. Schulam, S. Burger, D. Ding, Y. Wang, and F. Metze, “Robust audio-codebooks for large-scale event detection in consumer videos,” in Proc. Interspeech, 2929-2933 (2013).
  14. V. Carletti, P. Foggia, G. Percannella, A. Saggese, N. Strisciuglio, and M. Vento, “Audio surveillance using a bag of aural words classifier,” in Proc. IEEE Inter. Conf. Ad. Video and Sig. Surveillance, 81-86 (2013).
  15. T. George and P. Cook, “Musical genre classification of audio signals,” IEEE Trans. Speech and Audio Process. **10**, 5, 293-302 (2002).
  16. The HTK book Version 3.4, Cambridge University Engineering Department, (2009).

## 저자 약력

### ▶ 박 상 욱(Sang-wook Park)



2012년 2월: 중앙대학교 전자전기공학부  
(전자전기공학사)  
2012년 3월~현재: 고려대학교 전기전자  
전파공학부 석박사 통합

### ▶ 최 우 현(Woo-hyun Choi)



2009년 2월: 고려대학교 전자 및 정보공학  
부(학사)  
2009년 3월~현재: 고려대학교 전기전자  
전파공학부 석박사 통합

### ▶ 고 한 석(Hanseok Ko)



1982년 5월: 미국 카네기 멜론 대학교 전  
기공학 (공학사)  
1986년 5월: 미국 메릴랜드 대학교 시스템  
공학(공학석사)  
1988년 5월: 미국 존스 홉킨스 대학교 전  
기공학 (공학석사)  
1992년 5월: 미국 카톨릭 대학교 전기공학  
(공학박사)  
1995년 3월~현재: 고려대학교 전기전자전  
파공학부 교수