Journal of Korea Multimedia Society Vol. 16, No. 11, November 2013(pp. 1272-1280)
http://dx.doi.org/10.9717/kmms.2013.16.11.1272

PTZ 카메라 감시를 위한 실시간 위험 소리 검출 및 음원 방향 추정 소리 감시 시스템

응웬비엣쿡[†], 강호석[††], 정선태[†††], 조성원[††††]

요 약

본 논문에서는 실시간으로 위험한 소리를 인식하고 그 방향을 파악하여 이를 통해 PTZ Camera가 위험한 소리 방향으로 회전하여 해당 지역 영상을 획득하여 전송할 수 있도록 지원하는 소리 감시 시스템을 제안한다. 제안 소리 감시 시스템은 적응 혼합 가우시안 모델(AGMM)을 사용하여 일상적인 배경 소리와는 비정상적인 소리를 전경 소리로 검출하고, AGMM 모델로 미리 학습된 전경 소리들 중의 하나로 분류한다. 분류된 소리가 위험한 소리에 속하는 경우, Dual delay-line 방법에 기반을 둔 음원 방향 추정 기법을 사용하여 그 방향을 파악한다. 최종적으로 방향 정보를 사용하여 PTZ 카메라를 조절하여 그 방향 지역의 해당 영상을 획득하고 전송될 수 있도록 지원한다. 제안하는 소리 감시 시스템은 전경 위험 소리들을 안정적으로 검출하고, 79%의 정확도로 위험소리들을 분류하고, 작은 오차범위 이내 음원 방향 추정 성능을 나타냄을 실험 결과를 통해 확인하였다.

A Real-time Audio Surveillance System Detecting and Localizing Dangerous Sounds for PTZ Camera Surveillance

Viet Quoc Nguyen[†], HoSeok Kang[††], Sun-Tae Chung[†††], Seongwon Cho[††††]

ABSTRACT

In this paper, we propose an audio surveillance system which can detect and localize dangerous sounds in real-time. The location information about dangerous sounds can render a PTZ camera to be directed so as to catch a snapshot image about the dangerous sound source area and send it to clients instantly. The proposed audio surveillance system firstly detects foreground sounds based on adaptive Gaussian mixture background sound model, and classifies it into one of pre-trained classes of foreground dangerous sounds. For detected dangerous sounds, a sound source localization algorithm based on Dual delay-line algorithm is applied to localize the sound sources. Finally, the proposed system renders a PTZ camera to be oriented towards the dangerous sound source region, and take a snapshot against over the sound source region. Experiment results show that the proposed system can detect foreground dangerous sounds stably and classifies the detected foreground dangerous sounds into correct classes with a precision of 79% while the sound source localization can estimate orientation of the sound source with acceptably small error.

Key words: Audio Surveillance(오디오 감시), Dangerous Sound Detection(위험소리 검출), Sound Source Localization(음원 위치 검출), PTZ Camera(PTZ 카메라)

※ 교신저자(Corresponding Author) : 조성원, 주소 : 서울특별시 마포구 와우산로94(121-791), 전화 : 02) 3141-9540, FAX : 02) 320-1193, E-mail : swcho@hongik.ac.kr
접수일 : 2013년 9월 11일, 수정일 : 없음
완료일 : 2013년 10월 8일
[†] 준회원, 숭실대학교 정보통신전자공학부
(E-mail : nvquoc.uit@gmail.com)
[††] 정회원, 숭실대학교 정보통신전자공학부
(E-mail : dosanim@ssu.ac.kr)
[†††] 정회원, 숭실대학교 정보통신전자공학부
(E-mail : cst@ssu.ac.kr)
[††††] 정회원, 홍익대학교 전자전기공학부
※ 본 논문은 2013년 한국멀티미디어학회 춘계 학술대회 우수논문으로 추천되어, 학술지 게재로 추천되어 학술지 논문을 위해 그 내용을 보완하고 확장하여 다시 작성한 논문입니다.
※ 본 연구는 한국연구재단 기초 연구사업[2012 R1A1A2006883] 및 숭실대학교 교내연구비 지원을 받아 수행되었습니다. 연구비 지원에 감사드립니다.

1. 서 론

한정된 시야 범위 내의 영상만 감시 가능한 고정형 카메라의 단점을 극복하기 위하여 최근에는 광역 시야범위를 감시 할 수 있는 팬-틸트-줌 카메라(PTZ Camera)가 영상 감시에 많이 사용되고 있다. 지능형 영상 감시 시스템에서는 배경의 차이를 이용하여 전경 객체를 검출하고 배경을 갱신하거나[1], 전경 이동 객체를 검출하고 추적하는 방법을 사용하여 PTZ Camera를 제어할 수 있다[2]. 그러나 영상 감시의 경우 카메라의 시야범위 영역만을 감시할 수 있으므로, 시야 영역을 벗어나거나 어두운 곳에서 발생하는 상황은 감시할 수 없다. 만일 카메라를 필요한 시점에 원하는 방향으로 향하게 할 수 있다면 큰 도움이 될 것이다. 이를 위하여 카메라 주변의 소리를 분석하여 특정한 소리를 검출하고[3], 그 음원의 방향을 파악하여 카메라의 시야 범위를 변경하여 영상을 특정한 경우에만 얻어 온다면, 전원이나 전송대역도 절약하면서 카메라의 제약 조건도 극복할 수 있다. 특정한 소리를 검출하고 그 방향을 파악하는 문제는 이동식 로봇 분야나 카메라 감시 분야에서 활발히 진행되고 있다.

본 논문에서는 위험한 소리를 검출하고 방향을 파악하여 그 음원이 있는 지역의 영상을 얻어 올 수 있도록 PTZ Camera를 지능적으로 제어하는 소리 감시 시스템을 제안한다. 제안하는 감시 시스템은 크게 네 가지 모듈 {전경 소리 검출 (**FSD** : **F**oreground **S**ound **D**etection) 모듈, 소리 분류 (**SC** : **S**ound **C**lassification) 모듈, 음원 방향 추정 (**SSL** : **S**ound **S**ource **L**ocalization) 모듈과 PTZ Camera 제어 (**PCC** : **P**TZ **C**amera **C**ontrol) 모듈} 들로 구성 된다 우선 전경 소리 검출 (FSD) 모듈은 적응 혼합 가우시안 모델 (Adaptive Gaussian Mixture Model)을 사용하여 건물 내의 일상적인 소리를 배경 소리로 모델한 후 일상적이 아닌 소리를 검출한다[4]. 그 후 소리 분류 (SC) 모듈은 검출된 소리들을, GMM 모델을 사용하여 미리 학습된 유형 중의 하나로 분류하고[5], 그 유형이 위험한 소리에 속하는 지를 판정한다. 만일 위험한 소리에 속한다면, 음원 방향 추정 (SSL) 모듈이 Dual delay-line algorithm[6] 에 기반을 둔 음원 방향 추정 기법을 사용하여 그 방향을 파악한다. 마지막으로 PTZ Camera 제어 (PCC) 모듈이 PTZ Camera를 위험한 음원의 방향으로 향하게 하여 영상을 획득하여 전송하도록 한다.

실제로 사용되는 조그만 크기의 PTZ Camera 에 장착되어 실시간에 실행되기 위해서는 음원 방향 추정 (SSL) 모듈의 크기도 작고 알고리즘의 복잡도도 낮아야만 한다. 보통 흔히 사용되는 상관관계 기반의 PHAT 방식은, 마이크와 마이크 사이의 간격이 0.5m 이고 96 KHz의 높은 sampling rate를 사용하거나[7], 0.3m 간격에 44.1 KHz의 sampling rate를 사용해야 하기 때문에[8], 크기도 커지고, 복잡도도 높아져서 실시간 임베디드 환경에서 사용되기는 어렵다. 제안하는 시스템에서는 Dual line-delay 방법을 구현하여, 0.1m의 짧은 간격과 16 KHz의 낮은 sampling rate 를 사용하여, 계산 시간을 줄이고 소형 감시 시스템에서도 사용할 수 있도록 하였다. 이미 전경 소리 검출 (FSD), 소리 분류 (SC), 음원 방향 추정 (SSL) 분야에서 많은 연구들이 각각 진행되고 있지만, 본 논문은 이들을 통합하고 최적화하여 작은 PTZ Camera 에서 실시간에 실행 가능하도록 한 점이 다르다고 할 수 있다.

본 논문의 구성은 다음과 같다. 제2절에서는 제안하는 소리 감시 시스템 구성에 대해 소개하며, 제3절 및 4절에서는 본 논문의 전경 소리 검출 방법 및 음원 방향 추정 방법에 대해 자세히 기술한다. 실험 결과가 제5절에서 설명되며, 마지막으로 제6절에서는 결론이 기술된다.

2. 제안하는 소리 감시 시스템

제안하는 소리 감시 시스템은 그림 1과 같이, 전경 소리 검출 모듈, 소리 분류 모듈, 음원 방향 추정 모듈과 PZT Camera 제어 모듈의 4가지 모듈로 구성된다.

우선, 전경 소리 검출 모듈은 다음 세 가지 기능을 제공한다.

- **특징 추출(Feature extraction) 기능 :** 이벤트가 검출된 시계열을 생성하기 위해 소리 데이터에서 저수준의 오디오 특징을 추출한다. 오디오 특징으로는 좋은 소리 특징 벡터를 제공한다고 알려진 MFCC (Mel Frequency Cepstral Coefficients)[5,9]를 사용한다.
- **전경 소리 검출(Foreground Sound Detection) 기**

능 : 소리 데이터의 특징 벡터를 분석하여 미리 훈련된 적응 혼합 가우시안 배경 소리 모델과 비교하여 문턱값을 넘는 지를 확인한다. 새로 들어오는 소리가 전경 소리가 아닌 경우에는 아래 방법을 사용하여 적응 혼합 가우시안 배경 소리 모델을 갱신한다.

- **배경 소리 모델 갱신 (Updating new information for Background Model) 기능 :** 여기서는 적응 혼합 가우시안 모델용 증가적 학습 알고리즘(Incremental Learning Algorithm for Adaptive Gaussian Mixture Model)을 사용하여 배경 소리 모델을 갱신한다[4]. 전경 소리 검출 기능에서 배경소리에 속한다고 결정된 저수준 오디오 특징을 배경 소리 모델로 갱신하는 것이다.

전경 소리 검출 모듈은 새로운 소리가 들어오는 즉시 비일상적인 전경소리를 검출할 수 있도록 구현되었다. 그리고 증가적 학습 알고리즘은 환경이 바뀔 때 배경 모델을 적응적으로 갱신하도록 해 준다[10].

두 번째, 소리 분류 모듈은 전경 소리 클래스 데이터베이스와 전경 소리 분류 기능의 두 가지로 구성된다.

- **전경 소리 클래스 데이터베이스(Database of Foreground Classes) :** 배경 소리 모델과 마찬가지로, 전경 소리 클래스 데이터베이스는 각 소리 클래스의 특징 벡터의 분포를 나타내는 혼합 가우시안 모델을 저장한다. 본 논문에서는 6가지 유형 {박수소리, 유리 깨지는 소리, 울음소리, 말소리, 비명소리, 걸음소리}들을 전경소리로 분류하고, 실제 사용될 환경에서 녹음한 소리로 만든 데이터베이스들을 훈련하여 모델을 만들었다. 각각의 데이터베이스의 클래스에는 각각의 위험수준을 설정해 놓고, 이와는 독립적으로 실행할 때 위험그룹과 안전그룹을 결정한다.
- **전경 소리 분류 기능(Classifying Foreground Sound) :** 이 기능은 전경 소리가 속하는 클래스를 데이터베이스에서 결정할 수 있도록 추론한다. 미리 정의되고 훈련된 전경소리의 저장된 모델 데이터베이스로부터, 새로운 소리가 전경 소리가 될 가능성을 계산하여 클래스를 분류하고, 각 클래스에 설정된 위험수준을 바탕으로 위험한지 여부를 결정한다.

세 번째, 음원 방향 추정 모듈은 위에서 검출된 위험한 소리가 어디로부터 오는 지를 추정하는 역할을 한다. 이를 위하여 다음 두 가지 기능을 제공한다.

- **소리 도착시간 차이 계산(Calculating TDOA between microphones) :** TDOA (도착시간 차

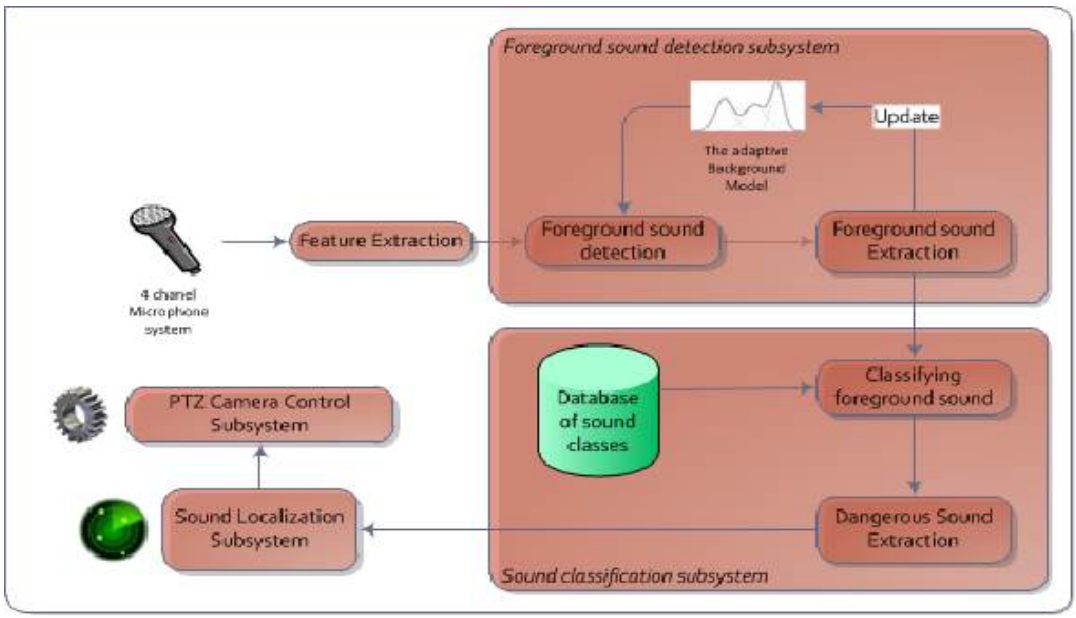

그림 1. 제안 소리 감시 시스템 구성

이, Time Difference Of Arrival)는 음원 방향 추정에서 가장 많이 사용되는 방법 중의 하나이다. 본 논문에서는 Dual delay-line 방법을 사용하여 마이크와 마이크 사이의 TDOA를 계산한다. TDOA를 사용하면 음원과 여러 마이크 사이의 거리 차이들을 쉽게 구할 수 있다.

- **음원 방향 계산(Calculating Position of Sound Source)** : 마이크와 음원 사이의 거리 차이들을 사용하여 기하학적으로 계산하면 음원이 어디에서 온 지를 정확하게 계산할 수 있다.

위험한 소리가 검출되고 그 방향이 계산되면, PTZ Camera 제어모듈을 사용하여 카메라가 그 방향을 향하도록 조정하고, 영상을 얻어 와서 영상감시를 계속하도록 한다.

3. 전경 소리 검출

영상 감시 시스템에서는 PZT Camera 시야 영역 내에서 어떤 움직임이 감지되면 카메라를 제어하여 그 부분을 자세히 감시할 수 있다. 본 시스템에서는 소리 감시 시스템을 추가하여, 평상시에 들리는 배경 소리와 다른 소리를 검출한다. 예를 들어 건물 내의 조용한 환경이라면, 장비나 팬에서 들리는 약간의 소음 등을 배경 소리라고 할 수 있는데, 갑자기 누군가의 비명소리가 들린다면 그 소리를 검출하는 것이다. 이때 비명소리가 들리는 즉시 반응할 수 있는 것이 중요하다. 이를 위하여 일상적인 배경 소리 모델을 미리 만든다.

배경 소리 모델의 저수준 특징들로 구성된 시계열 $\{O_1O_2O_3\cdots O_N\}$을 가정해 보자. 여기서 O_i는 시간 i 때의 저수준 특징 벡터이다. 이를 혼합 가우시안 모델로 훈련시켜서 예측한 모델을 배경 *소리 모델 G_b라고 하자. 적응적 배경 소리 모델은 다음처럼 예측될 수 있다.

3.1 배경 소리 모델의 예측

저수준 특징 벡터들인 $\{O_1O_2O_3\cdots O_N\}$을 사용하여 훈련세트 T_N를 만들어 보자. N은 T_N의 배열 크기이고 Kmax는 적응적 배경 소리 모델의 가우시안 구성요소의 개수 이다. 혼합 가우시안 모델의 매개 변수 세트 $\{\mu^{(i)}\}_{i=1}^{Kmax}, \{\pi^{(i)}\}_{i=1}^{Kmax}, \{R^{(i)}\}_{i=1}^{Kmax}$ 의 K번째 구성요소인 $\theta^{(K)}$는 각각 평균, 혼합계수, 분산으로 구성되고, 이는 $\theta^{(K)}=\{\mu^{(K)},\pi^{(K)},R^{(K)}\}$ 로 나타낼 수 있다.

특징 벡터 배열 T_N의 확률의 log를 취하면 다음 식과 같다.

$$\log(p(T_N|\theta^{(K)}))=\sum_{n=1}^{N}\log\sum_{k=1}^{K}(N(O_n|\mu_k^{(K)},R_k^{(K)})\pi_k^{(K)}) \quad (1)$$

우리의 목표는 K를 예측하고, 이에 해당하는 각 $\theta^{(K)}$를 구하는 것이다. 각 K에 대하여 θ를 예측할 때 기대값 최대화 방법이 많이 사용된다. 그러나 K를 찾는 알고리즘이 아직 없기 때문에, 가장 큰 값부터 점점 줄여 가면서 벌점을 계산하는 방법인 MDL 예측기[11]를 사용한다.

K를 줄여 나갈 때 마다 배경 소리 모델의 구성요소의 개수도 K개에서 K-1개로 줄여야 한다. 우리는 가장 근접한 두 개의 구성요소를 하나로 합병하였다. 만일 l과 r이 가장 가깝다면, 합병된 하나의 구성요소 (l,r)는 다음과 같이 만들 수 있다.

$$\pi_{(l,r)}=\pi_l+\pi_r \quad (2)$$

$$\mu_{(l,r)}=\frac{\pi_l\mu_l+\pi_r\mu_r}{\pi_l+\pi_r} \quad (3)$$

$$R_{(l,r)}=\frac{\pi_l(R_l+(\mu_l-\mu_{(l,r)})(\mu_l-\mu_{(l,r)})^T)}{\pi_l+\pi_r}+\frac{\pi_r(R_r+(\mu_r-\mu_{(l,r)})(\mu_r-\mu_{(l,r)})^T)}{\pi_l+\pi_r} \quad (4)$$

여기서 $\pi_l,\pi_r,R_l,R_r,\mu_l,\mu_r$는 θ의 매개 변수들이다.

각각의 다른 환경의 각각의 훈련 세트마다 예측 알고리즘을 사용하여 다른 배경모델을 만든다. 이 알고리즘을 사용하여 훈련된 모든 배경 모델들을 배경 모델 세트 S_b에 미리 저장 해 놓는다. 다음 절에서는 S_b를 사용하여 배경 소리를 검출하는 방법을 설명한다.

3.2 전경 소리 검출 및 배경 소리 모델 갱신

시스템이 시작했을 때 녹음된 소리 데이터로부터 추출한 저수준 특징 벡터를 T_N=$\{O_1O_2O_3\cdots O_N\}$이라고 하면, 일단은 시스템이 실행되는 환경의 기본적인 배경 소리 모델로 T_N을 사용할 수 있다.

만일 θ_b가 시스템이 실행되는 환경의 가장 적합한 배경소리 모델 이라면, 우리는 배경 모델 세트 S_b로부터 다음과 같이 θ_b 를 선택할 수 있다.

$$\theta_b = \arg max_\theta \, p(O_1, O_2, O_3, ..., O_N|\theta) \quad (5)$$

비록 θ_b가 가장 적합한 배경 소리 모델이라고 해도, 아직까지는 현재 환경의 모든 배경 소리 정보를 모델하지는 못했기 때문에, 증가적인 학습 알고리즘을 사용하여 T_N의 새로운 정보를 θ_b에 갱신한다.

시스템이 시작 되어 처음 훈련세트 $\{O_1 O_2 O_3 ... O_N\}$로부터 예측된 배경 소리 모델이 있을 때, 그 후 마이크로부터 들어온 새로운 소리에서 특징 벡터 $\{O_1 O_2 O_3 ... O_M\}$을 추출한다. 만일 이 소리가 전경 소리라면 다음과 같이 예측된 차이 d가 클 것이다.

$$d = \log(p(O_1, O_2, O_3, ..., O_M|\theta) - \log(p(O_1, O_2, O_3, ..., O_N|\theta) \quad (6)$$

반면에, 새로운 소리가 배경 소리라면 d가 크지 않을 것이다. 이 경우 증가적 학습 알고리즘을 사용하여 새로운 정보를 배경 소리 모델에 갱신한다.

다음 절에서는 음원 방향 추정 모듈을 설명한다.

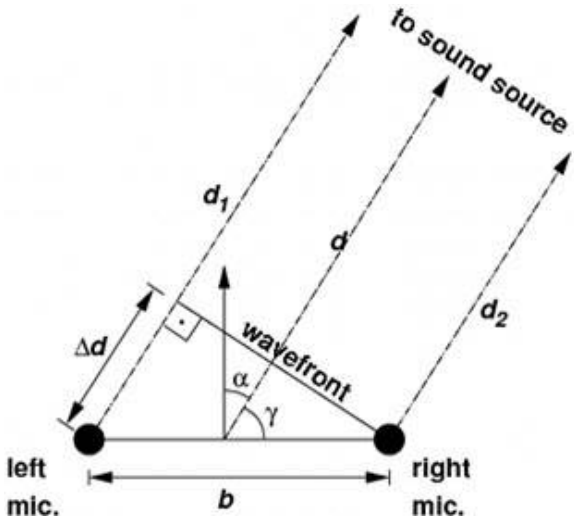

그림 2. 음원 방향 추정 모듈

4. 음원 방향 추정

음원의 방향을 추정하는 방법은 많이 있지만, PTZ Camera에서 사용되기 위해서는 임베디드 시스템에서 실시간으로 처리하기 위하여 계산시간이 적어야 한다. 하지만 많이 사용되는 상관관계 기반의 PHAT 방법[7] 은 96 KHz의 높은 sampling rate 와 마이크 사이의 간격이 0.5m 나 되어야 하기 때문에 높은 복잡도의 계산이 필요하다.

따라서, 제안하는 시스템에서는 Dual delay-line 방법[6]을 사용하여 도착 방향을 계산한다. 그림 2의 음원 방향 추정 모듈은 다음과 같이 도착 방향을 계산한다.

4.1 2D 음원 방향 추정

동시에 한 쌍의 마이크에서 얻어진, 멀리 떨어진 방향으로부터 온 소리 신호 $s_1(k), s_2(k)$ 로부터 T개의 소리 샘플을 얻은 후, 각각 이산 푸리에 변환을 하면 다음과 같은 스펙트럼을 얻을 수 있다.

$$S_1(m) = F(s_1(k)) \quad (7)$$
$$S_2(m) = F(s_2(k))$$

여기서 푸리에 계수 m에 해당하는 주파수 f_m (in Hz)는 다음과 같다.

$$f_m = \frac{m}{T} \times Samplerate \quad (8)$$

만일 $2 \times \triangle t$ 가 $s_1(k), s_2(k)$ 사이의 시간 차이라고 하면, $s_1(k)$과 $s_2(k)$는 같은 음원에서 왔으므로 다음과 같은 결과를 얻는다.

$$S_1(m) e^{j2\pi f_m \triangle t} = S_2(m) e^{-j2\pi f_m \triangle t} \quad (9)$$

그러면 그림 2에서 두 개의 소리 신호 $s_1(k), s_2(k)$ 사이의 도착 각도 α를 다음과 같이 예측할 수 있다.

$$\alpha = \arg min_{\triangle t} \sum_{m=0}^{\frac{T}{2}-1} |S_1(m) e^{j2\pi f_m \triangle t} - S_2(m) e^{-j2\pi f_m \triangle t}| \quad (10)$$

α가 90° 일 때 $\triangle t$가 최대값이 되고, α가 -90° 일 때 $\triangle t$가 최소값이 되며, $\triangle t$ 와 α 사이의 관계는 다음과 같다.

$$2 \triangle t = \frac{b \sin(\alpha)}{v_s} \quad (11)$$

여기서 b는 2개의 마이크 사이의 거리 이고, v_s는 속도이다.

식 (10)과 같은 비선형식을 푸는 것은 어렵기 때문에 예측 방법을 생각 해 보자. -90°에서 +90° 사이의 방위각 (Azimuth) 공간을 같은 크기의 I sector로 나누면, i 번째sector를 사용하여 다음과 같이 α를 예측할 수 있다.

$$\alpha = \frac{i}{I-1}\pi - \frac{\pi}{2} \quad (12)$$

그러면 식 (10)은 다음과 같이 된다.

$$\alpha = \arg\min_i \sum_{m=0}^{\frac{T}{2}-1} \left| S_1(m) e^{j2\pi f_m \frac{b\sin(\frac{i}{I-1}\pi - \frac{\pi}{2})}{2v_s}} - S_2(m) e^{-j2\pi f_m \frac{b\sin(\frac{i}{I-1}\pi - \frac{\pi}{2})}{2v_s}} \right| \quad (13)$$

식 (13)을 사용하면 0에서 I-1 까지의 i 값을 사용하여 α를 예측할 수 있다. 그 정밀도는 I 값에 비례하여 높아지지만, I 값이 너무 크면 많은 계산 시간이 필요하게 된다.

4.2 3D 음원 방향 추정

4개의 마이크 1, 2, 3 과 4에서 얻은 소리 신호를 $s_1(t), s_2(t), s_3(t)$ 과 $s_4(t)$ 라고 하자. 4개의 마이크는 그림 3처럼 PTZ Camera에 부착한다. 음원의 방향을 결정하기 위하여 카메라 렌즈 모듈을 중심에 두는 구면 좌표계를 사용한다. 마이크 1과 2는 수평면에 부착하고, 마이크 3과 4는 수직면에 부착한다. 음원의 방향은 방위각 Φ와 고도 ω를 사용하여 (Φ, ω) 로 나타 낼 수 있다. 방위각 Φ는 식 (13)에 $s_1(t), s_2(t)$를 대입하여 구하고, 고도 ω는 역시 식 (13)에 $s_3(t), s_4(t)$를 대입하여 예측하면 된다.

5. 실험 결과

본 논문에서는 3가지 다른 환경 (사무실, 엘리베이터, 복도)에서 PTZ Camera에 부착한 마이크들로부터 직접 녹음한 3개의 소리 파일을 사용하여 전경 소리 검출 모듈을 테스트 하였다. 각각의 소리 파일은 20분에서 35분간 녹음하였고, 저수준 특징으로는 MFCC를 사용 하고, 각 frame은 0.06초 길이에 0.02초씩 겹치게 하였다. 바람 부는 환경, 사무실, 복도, 길거리, 기타 등의 환경에서 녹음한 배경 소리를 미리 훈련하여 배경 소리 모델 세트를 구성하였는데, 대부분의 테스트에서 사무실 환경이 가장 적합한 배

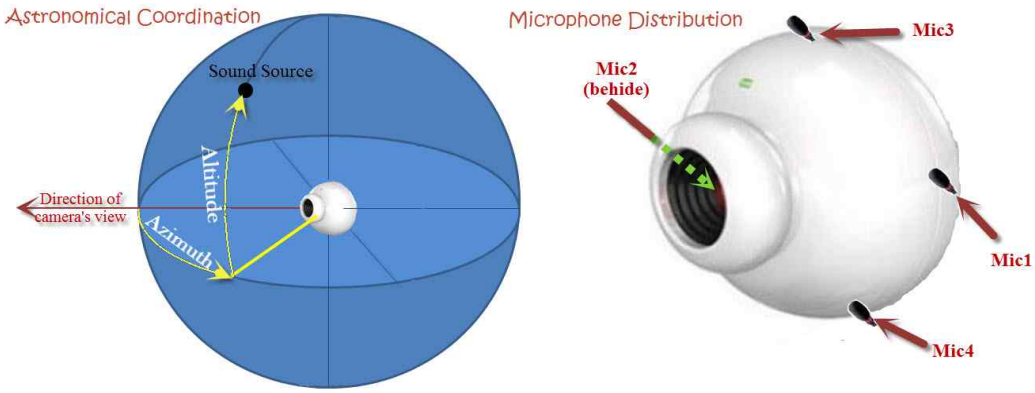

그림 3. 구면 좌표계와 PTZ Camera, 4개의 마이크 부착 방법

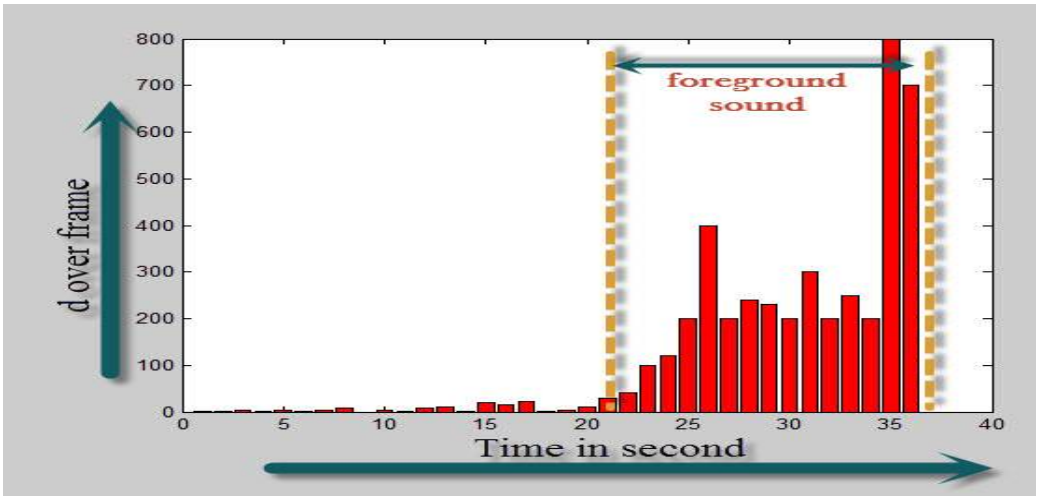

그림 4. 사무실 환경의 전경 소리 검출 결과

경 모델로 선택 되었다.

사무실에서 전경 소리를 검출하는 실험 결과를 그림 4에서 보여 주는데, y 축은 식 (6)에서 계산한 사무실의 배경 소리와 녹음 된 소리 사이의 d를 나타내고, x 축은 경과 시간을 나타낸다.

그림 4에서 보면, 전경 소리의 경우, 제3절에서 설명한 바와 같이 차이가 큼을 알 수 있다. 즉, d 의 차이가 큰 경우에 쉽게 전경 소리로 검출되며 따라서 전경 소리 검출은 매우 안정적으로 수행됨을 다양한 실험을 통해 확인하였다.

소리 분류 모듈을 위하여 6가지 전경 소리 유형{박수소리, 유리 깨지는 소리, 울음소리, 말소리, 비명소리, 걸음소리}들을 미리 훈련하였다. 각각의 유형은 50에서 100개의 소리 샘플로 훈련하였고, 데이터베이스의 길이로는 450초를 사용하였다. 소리 분류 모듈을 실험할 때는, 67%의 데이터베이스로 훈련하고 33%의 데이터베이스를 사용하여 테스트 하였는데, 그림 5에 보이는 것처럼 평균 79%의 올바른 인식율을 보였다. 일반적으로 사용되는 가우시안 혼합 모델을 사용한 결과[12]인 약 45%에 비해서 제안하는 시스템이 실시간에 더 좋은 결과를 보여 주었다.

음원 방향 추정 모듈은 2개의 마이크를 PTZ Camera에 10.6 cm 간격으로 부착하고, 16 KHz 로 샘플하여 실험 하였다. 소리의 속도는 대략 340 m/s 이지만 환경의 온도에 따라 변한다. 이러한 구성으로 테스트 한 결과 상당히 정확한 음원의 도착 각도를 계산할 수 있었고, 표 1에 보인 대로 1° 미만의 오차를 기록 하였다. 이는 PHAT에 기반한 방법[7,13] 이나 일반적인 상관관계에 기반한 방법[8]에 비하여 제안하는 시스템이 실시간 임베디드 감시 시스템에 더 적합함을 보여준다.

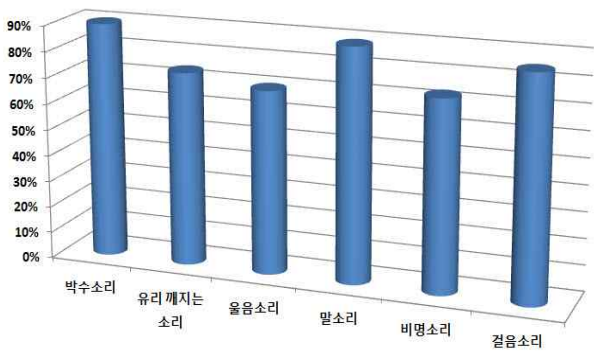

그림 5. AGMM 기반의 전경 소리 분류 결과

표 1. 음원의 도착 각도 계산 결과와 절대 오차

실제 각도 (°)	음원 추정 결과 (°)	절대 오차
-90	-90	0
-75	-76	1
-60	-60.9	0.9
-45	-44.3	0.7
-30	-30.4	0.4
-15	-14.4	0.6
0	0.2	0.2
15	15.5	0.5
30	29.5	0.5
45	44.6	0.4
60	79	1
75	76.1	1.1
90	90	0

6. 결 론

본 논문에서는 실시간에 위험한 소리를 검출하고 음원의 방향을 추정하여 PTZ Camera를 조절하는 소리 감시 시스템을 설명하였다. 적응 혼합 가우시안 모델용 증가적 학습 알고리즘을 사용하여 환경의 변화에 따라 배경 소리 모델을 갱신하여, 일반적인 가우시안 혼합 모델을 사용한 결과인 약 45% [12]에 비해서 전경 소리를 약 79% 검출할 수 있었다. 일반적 상관관계를 사용한 기존의 연구[7,8,13] 대신에 Dual delay-line 기반[6]의 방법을 사용하여, 마이크 사이의 간격을 0.3~0.5m에서 0.1m로 줄이고, 44.1~96 KHz의 높은 샘플링 레이트 대신에 16 KHz의 낮은 샘플링 레이트를 사용하여 계산 시간을 많이 줄일 수 있었다. 또한 위험 소리 검출 및 음원 방향 추정 기능을 통합하고 최적화 하여 실시간에 소형 PTZ 카메라를 컨트롤 하는 임베디드 시스템에서 구현 가능 하도록 하였다.

향후에는 실제 환경 하에서 엄격한 테스트를 거친 후 상용화 할 수 있도록 개선할 예정이다.

참 고 문 헌

[1] S. Jung and M. Kim, “Techniques for Background Updating under PTZ Camera Based Surveillance,” *Journal of Korea Multimedia*

Society, Vol. 12, No. 12, pp. 1745-1754, 2009.

[2] A. Bevilacqua and P. Azzari, "High-quality Real Time Motion Detection Using PTZ Cameras," *Proc. IEEE Int. Conf. on Video and Signal Based Surveillance*, pp. 23-28, 2006.

[3] K. Kalgaonkar, P. Smaragdis, and B. Raj, "Sensor and Data Systems, Audio-Assisted Cameras and Acoustic Doppler Sensors," *IEEE Proc. On Computer Vision and Pattern Recognition*, pp. 1-2, 2007.

[4] M. Cristani, M. Bicego, and V. Murino, "On-line Adaptive Background Modelling for Audio Surveillance," *Proc. 17th Int. Conf. on Pattern Recognition*, pp. 399-402, 2004.

[5] R. Radhakrishnan, A. Divakaran, and P. Smaragdis, "Audio Analysis for Surveillance Applications," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 158-161, 2005.

[6] C. Liu, B.C. Wheeler, W.D. O'Brien, Jr, R.C. Bilger, C.R. Lansing, and A.S. Feng, "Localization of Multiple Sound Sources with Two Microphones," *Journal of the Acoustical Society of America*, Vol. 108, No. 4, pp. 1888-1905, 2000.

[7] A. Pourmohammad and S.M. Ahadi, "Real Time High Accuracy 3-D PHAT-Based Sound Source Localization Using a Simple 4-Microphone Arrangement," *IEEE Systems Journal*, Vol. 6, No. 3, pp. 455-468, 2012

[8] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Scream and Gunshot Detection and Localization for Audio-Surveillance Systems," *Advanced Video and Signal Based Surveillance*, pp. 21-26, 2007.

[9] S.B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 28, No. 4, pp. 357-366, 1980.

[10] M. Song and H. Wang, "Highly Efficient Incremental Estimation of Gaussian Mixture Models for Online Data Stream Clustering," *Proc. of SPIE Conference on Intelligent Computing*, pp. 174-183, 2005.

[11] Z. Xiong, R. Radhakrishnan, A. Divakaran, and T.S. Huang, "Effective and Efficient Sports Highlights Extraction Using the Minimum Description Length Criterion in Selecting GMM Structures," *Proc. of ICME*, Vol. 3, pp. 1947-1950, 2004.

[12] M. Cowling and R. Sitte, "Comparison of Techniques for Environmental Sound Recognition," *Pattern Recognition Letters*, Vol. 24, No. 15, pp. 2895-2907, 2003.

[13] Viet Quoc Nguyen, 강호석, 정선태, 설태인, 조성원, "팬-틸트-줌 카메라를 위한 위험소리 검출 및 위치 추적," 한국멀티미디어학회 춘계학술발표대회 논문집, 제16권, 제1호, pp. 20-23, 2013.

응 웬 비 쿡

2012년 9월 베트남 호지민 국립대학교 정보통신대학 전산과 공학사
2012년 9월 ~ 현재 숭실대학교 대학원 정보통신전자공학과 석사과정

관심분야 : 신호처리, 오디오처리, 임베디드 시스템

강 호 석

1985년 2월 서울대학교 전기공학과 학사
1988년 5월 미국 University of Florida 전기전자공학 석사
1994년 12월 미국 Purdue University 전기전자공학 박사
현재 숭실대학교 정보통신전자공학부 교수
관심분야 : 영상처리, 컴퓨터 그래픽스, 임베디드 시스템

정 선 태

1983년 2월 서울대학교 전자공학과 학사
1986년 12월 미국 Michgan Univ. (Ann Arbor) 전자 및 컴퓨터공학과 석사
1990년 12월 미국 Michgan Univ. (Ann Arbor) 전자 및 컴퓨터공학과 박사
1991년 3월 ~ 현재 숭실대학교 정보통신전자공학부 교수
관심분야 : 영상처리, 컴퓨터 비전, 임베디드 시스템

조 성 원

1982년 2월 서울대학교 전기공학과 학사
1987년 12월 미국 Purdue University 전기전자공학 석사
1992년 7월 미국 Purdue University 전기전자공학 박사
현재 홍익대 전자전기공학부 교수
관심분야 : 영상처리 및 인식, 지능시스템