

기술백서

지능형 음원 분석 기술

분류 및 방향탐지 기술백서 & 설치 가이드

2026년 2월 9일

Contents

1. 개요
2. 배경
3. AI 기반 오디오 분석 기술
 - 3.1. 음원 분류(Sound Classification) 기술
 - 3.2. 음원 방향 탐지(Sound Direction Detection) 기술
4. 최적의 성능을 위한 설치 및 환경 가이드
 - 4.1. 최적의 설치 위치 선정
 - 4.2. 효과적인 음원 감지 및 분류를 위한 환경 분석
 - 4.3. 음원 분류 dB 임계값 설정
 - 4.4. 음원 방향 보정 및 시스템 설정
 - 4.5. 복합 음향 환경 및 예외적 음원 활용 팁
5. 결론

1. 개요

본 백서는 한화비전의 AI 기반 오디오 분석 기술을 통해 공격성 소리(Agression)를 포함한 음원 분류, 방향 탐지를 통합한 멀티모달 보안 솔루션을 제시한다. 이 기술들은 상호 유기적으로 연동되어 통합적 상황 인지 능력을 극대화하며, 차세대 보안 시스템의 새로운 기준을 제시한다. 보안 시스템 구축자, IT 관리자, 통합 보안 담당자 및 영상 보안의 한계를 넘어 소리 기반 조기 위협 탐지를 원하는 전문가들에게 유용하게 사용될 것으로 기대된다.

본 백서는 한화비전의 딥러닝 기반 음원 분류 기술과 음원 방향 탐지 기술을 깊이 있게 다루고, 사용자 환경에 최적화된 활용 방안을 제시하여, 단순 감지에서 상황 인지로의 진화와 더불어, 범죄 예방을 위한 사전 도구의 역할을 한다.

2. 배경

보이지 않는 위협 속에서 소리는 종종 간과되는 강력한 감시 도구 중 하나다. 기존 영상 보안 시스템이 무엇이 일어나고 있는지 시각적으로 포착하는 데 집중했다면, 오늘날의 보안 환경은 소리 이벤트의 유형 및 발생 지점까지 인지하는 단계로 진화했다. 공공 안전과 자산 보호의 경계가 확장됨에 따라, 오디오 분석 기술은 단순한 보조 수단을 넘어 범죄 예방 및 신속한 상황 대처에 기여할 잠재력을 갖추게 되었다. 멀티모달 보안 수요 증가로, AI 오디오 분석이 범죄 예방의 필수 요소로 부상 중이다.

AI 가 성장함에 따라 단순히 소리를 녹음하고 듣는 수준을 넘어 비정상적인 소음(비명, 유리창 깨지는 소리 등)을 실시간으로 탐지하고 분석하는 기술이 보안시장에 적용되었다. 이는 영상 감지의 기술적인 사각지대를 해소할 수 있어 환경적인 제약이나, 감지 범위의 제약 또는 영상을 촬영할 수 없는 탈의실이나 화장실, 병원과 같은 장소에서 영상 감지의 기술적인 사각지대를 해소할 수 있다.

이는 소리 기반 분석을 통해 시각 정보만으로는 인지하기 어려운 이벤트를 선제적으로 감지하도록 설계되었으며, 감지된 음성 이벤트는 다른 장비나 시스템으로 연동되어 관제 요원이 상황을 보다 신속하게 인지하고 대응할 수 있도록 지원한다.

프라이버시 침해에 대한 논란이 커지고 있는 요즘, 영상 노출에 대한 거부감이 커지면서 영상 외 다양한 센서를 활용한 감지 기술, 특히 개인 식별 요소를 최소화한 AI 오디오 분석 기술이 새로운 대안으로 주목받고 있다.

3. AI 기반 오디오 분석 기술

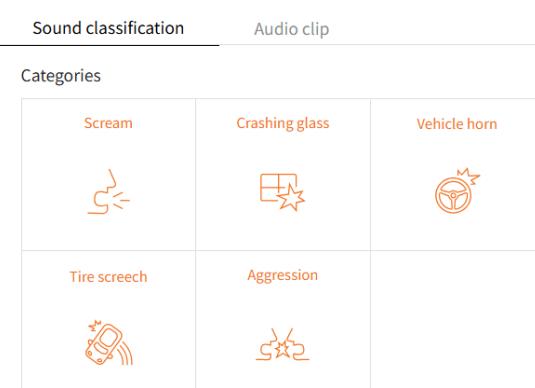
3.1. 음원 분류(Sound Classification) 기술

한화비전 음원 분류 기술은 딥러닝 기반의 CNN(Convolutional Neural Network) 컨볼루션 신경망을 핵심으로 한다. 이 기술은 소리라는 추상적인 정보를 CNN이 처리하기 용이한 시각적 형태인 스펙트로그램¹으로 변환하는 과정에서 시작한다. 이를 통해 비명, 유리 깨짐, 차량 경적, 타이어 미끄러짐, 공격성 소리 등의 소리 이벤트를 정확하게 식별하고 분류한다.

스펙트로그램은 오디오 파형의 '지문'과 같이, 특정 소리 유형이 가지는 고유한 패턴을 명확히 보여준다. CNN은 이러한 스펙트로그램 이미지에서 사람의 귀로는 구별하기 어려운 미묘한 음향적 특징과 패턴을 자동으로 학습하고 인식하는 데 탁월한 성능을 발휘한다. 이 중에서 공격성 소리(Aggression) 감지 기술은 오디오 데이터를 실시간으로 분석하여 사람이 공격적인 감정을 드러내며 고함을 지르거나, 이목을 끌기 위해 크게 소리칠 때 발생하는 음성을 인식한다. 이 기술은 단순히 소리의 크기(dB)를 측정하는 기존 방식과 달리, 딥러닝 기반 모델을 활용해 주파수 도메인에서 비언어적 음향 특징을 분석하는 것이 핵심이다. 이 기술은 발화 내용(언어적 특징)에 상관없이 에너지 변동, 주파수 분포, 스펙트럼 밀도 등과 같은 음향적 특징을 종합적으로 분석함으로써 공격성 소리(Aggression) 여부를 판단한다. 일상적인 상황에서도 사람의 공격적인 음성 특징을 효과적으로 탐지하여, 돌발 상황이나 위기 상황에 대한 선제적인 대응을 지원한다.

음원 감지 및 분류가 완료되면, 음원 스트림 전체에서 데이터 추출 단계를 자동으로 거친다. 이미 오디오 데이터 전처리를 통한 샘플링이 완료된 상태이므로, 이후 음원 분류 결과와 함께 분리된 음원은 오디오 클립 형태의 파일로 생성되며, 이는 메타데이터와 함께 제공되어 손쉽게 다운로드해 활용할 수 있다.

이러한 음원 분류 기술은 [한화비전 WiseSound App](#)을 지원하는 제품에서 제공 가능하다.



¹ 스펙트로그램(spectrogram)은 소리 신호를 시각적으로 표현한 그래프를 말한다.

3.2. 음원 방향 탐지(Sound Direction Detection) 기술

한화비전 음원 방향 탐지 기술은 사용자가 지정한 이벤트 음원이 인식될 때 해당 음원의 방향을 탐지해 사용자에게 상황을 알리고 신속한 대응을 지원한다. 이 기술은 물리적으로 분리된 다중 마이크 각각에 도달하는 음원 신호의 시간차(TDoA, Time Difference of Arrival)를 측정해 방향을 추정한다.

TDoA 알고리즘은 음원 발생 후 각 마이크에 도달하는 시간의 위상차를 분석해 실제 마이크에 도달하는 거리를 추정하며, 이를 통해 음원 발생 위치의 각도를 계산한다. 그림 1과 같이, 다중 마이크 시스템은 여러 개의 마이크(MIC1, MIC2, MIC3, MIC4)를 원형으로 배치해 소리 발생 지점과 각 마이크 간의 거리 차이(d_1, d_2, d_3, d_4)를 파악한다. 이 거리 차이를 바탕으로 소리가 각 마이크에 도달하는 시간차를 계산하는 것이 TDoA 알고리즘의 핵심이다.

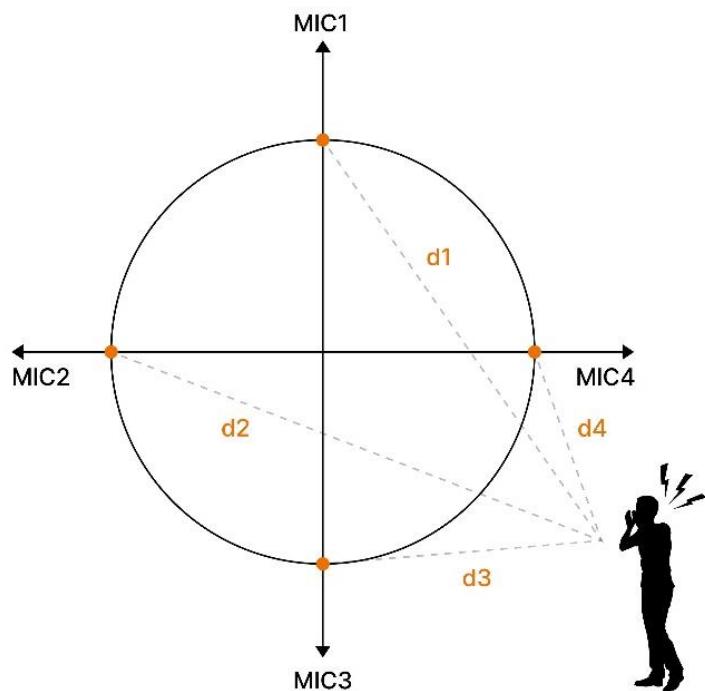


그림 1: 다중 마이크를 활용한 음원 방향 탐지

그림 2는 TDoA 알고리즘에서 두 마이크(갈색, 파란색 파형)에 소리 신호가 도달하는 시간차(τ_{ij})를 시각적으로 보여준다. 이처럼 소리 파형의 도달 시간 차이를 정밀하게 측정함으로써, 시스템은 소리 발생 지점의 정확한 방향을 역산할 수 있다.

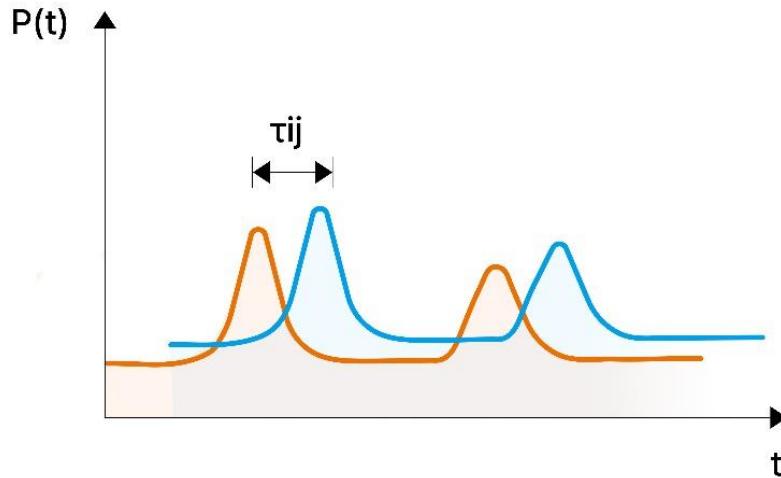


그림 2: TDOP 알고리즘의 시간차 측정 개념

방향 탐지 과정은 크게 4 단계로 구분된다.

1. **신호 수집 단계**: 다중 마이크를 통해 음원 신호를 동시에 수집한다.
2. **신호 처리 단계**: 수집된 음원 신호를 신호 알고리즘을 통해 분석한다.
3. **방향 추정 단계**: 처리된 신호를 기반으로 소리의 발생 방향을 추정한다.
4. **결과 출력 단계**: 최종적으로 탐지된 방향을 방위각으로 표시해 출력한다.

이 음원 방향 탐지 기술은 다중 마이크를 지원하는 오디오 비콘(Audio Beacon)과 일부 Wisenet 9 SoC 가 탑재된 카메라 등 한화비전 특정 제품에서 제공된다.



그림 3: 오디오 비콘 SPS-A100M(왼쪽)과 Wisenet 9 탑재 카메라의 마이크 위치

4. 최적의 성능을 위한 설치 및 환경 가이드

한화비전 AI 오디오 솔루션의 성능은 설치 환경의 특성과 밀접하게 연관된다. 다음 사항을 적극적으로 고려해 시스템의 잠재력을 최대한 활용하고 안정적인 성능을 확보할 수 있다.

4.1. 최적의 설치 위치 선정

안정적인 음원 분류 및 방향 탐지 성능을 보장하는 최적의 설치 조건은 다음과 같다.

- **음원 분류:** 제품과 소리 발생 지점 사이의 거리가 최소 2m 이상일 때 가장 안정적으로 동작한다. 이 거리는 음원(예: 사람의 키높이)을 기준으로 한다. 만약 거리가 2m 이내로 너무 가까울 경우 박수 소리와 같은 소리가 과도하게 크게 인식되어 오탐이 발생할 수 있다. 실내 천장 설치는 음향 반향을 최소화하고, 넓은 공간의 소리를 균일하게 감지하는 데 효과적이므로, 음원 분류 기능에 최적의 설치 방식이다.

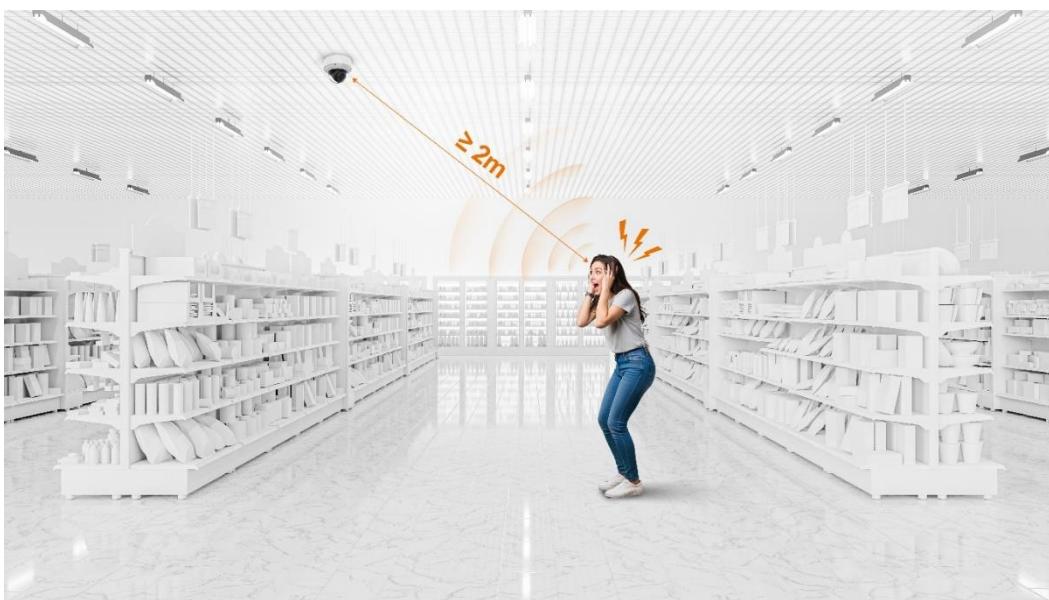


그림 4: 효과적인 음원 분류를 위한 설치 공간

- **직접적인 음원 경로 확보:** 음원 발생 지점과 제품 사이에 벽, 유리, 두꺼운 커튼 등의 물리적 장애물이 존재할 경우, 음원 신호가 감소하거나 왜곡될 수 있다. 시스템의 최대 성능을 위해서는 직접적인 음원 경로를 확보하는 것이 중요하다.

- **음원 방향 탐지:** 정확한 방향 탐지 성능을 위해서는 설치 장소의 최소 공간 크기가 가로 $\geq 6.0\text{m}$ X 세로 $\geq 6.0\text{m}$ 이상인 것을 권장한다. 이는 소리의 반향 및 잔향의 영향을 최소화하고 다중 마이크 간의 신호 분석에 필요한 충분한 공간을 확보하기 위함이다.



그림 5: 정확한 음원 방향 탐지를 위한 설치 공간

- **거리 및 입사각 준수:** 이벤트 음원 발생 지점과 제품 간의 거리 및 각도는 정확한 탐지 성능을 좌우하는 중요한 요소다. 이벤트 음원 발생 지점의 입사각이 20도 이상으로 너무 크거나, 제품과의 거리가 너무 가까울 경우 방향 탐지 정확도가 떨어질 수 있다. 다음 표는 제품 설치 높이에 따라 권장되는 최소 거리를 제시한다. 이 거리를 준수하면 더 신뢰할 수 있는 방향 탐지가 가능하다.

제품 설치 높이	방향 탐지 가능 최소 거리
2.3m	$\geq 2.2\text{m}$
2.5m	$\geq 2.7\text{m}$
2.7m	$\geq 3.3\text{m}$
2.9m	$\geq 3.8\text{m}$
3.1m	$\geq 4.4\text{m}$
3.3m	$\geq 4.9\text{m}$
3.5m	$\geq 5.5\text{m}$
3.8m	$\geq 6.3\text{m}$
4m	$\geq 6.9\text{m}$
5m	$\geq 9.6\text{m}$

표 1: 제품 설치 높이에 따른 음원 방향 탐지 최소 거리

4.2. 효과적인 음원 감지 및 분류를 위한 환경 분석

정확한 음원 감지 및 분류를 위해 다음과 같은 음압 조건과 주변 환경을 고려해야 한다.

음원 종류	한계 dB	한계 예측 거리
비명, 타이어 미끄러짐, Aggression	>65dB	2m ~ 16m
유리 깨짐	>60dB	2m ~ 7m
차량 경적	>60dB	2m ~ 20m

표 2: 음원 종류별 최소 음압 조건

예를 들어, 비명은 65dB 이상일 경우 정확한 분류 및 방향 탐지가 가능하며, 이벤트 음원의 dB 크기는 주변 배경 잡음보다 일정 수준 이상 커야 한다(권장: 15dB 이상). 또한, 주변 배경 잡음이 60dB 를 넘지 않는 장소에 설치해야 정확한 측정 및 분류가 가능하다. 이는 이벤트 음원과 잡음 간의 명확한 구분을 위함이다.

주변 잡음은 설치 환경에 따라 성능에 영향을 줄 수 있으므로, 다음과 같은 점들을 사전에 분석하면 좋다.

- **실외 환경:** 바람, 비, 번개 등 자연 소음이나 교통, 물체 충격, 자동차 반동 등 인공 소음이 있을 수 있다. 예측 불가능한 잡음이 많은 환경에서는 추가적인 환경 분석을 통해 최적의 설치 위치를 선정하는 것이 중요하다.
- **실내 환경:** 주변 재료(벽, 천장, 바닥) 및 방 크기에 따라 소리의 반향 및 잔향이 크게 발생할 수 있다. 풍선 터지는 소리, 무거운 상자를 떨어뜨리는 소리 등 특정 음원과 유사한 특징을 가진 소리는 잔향으로 인해 에너지가 축적되어 오탐지를 유발할 수 있으므로, 실내 공간의 음향 특성을 고려한 설치가 필요하다.

4.3. 음원 분류 dB 임계값 설정

음원 분류 기능의 성능을 최적화하기 위해 사용자의 환경에 맞춰 dB 임계값(Threshold)을 설정할 수 있다.

- 소음이 큰 환경에서는 임계값을 기본값보다 높게 설정해 불필요한 알람(오분류)을 줄일 수 있다.
- 이벤트 음원이 작은 환경에서는 임계값을 기본값보다 낮게 설정해 중요한 알람 누락(미분류)을 줄일 수 있다.
- 설치 환경의 배경 잡음 dB 를 확인한 후, 평균값보다 최소 15dB 이상의 임계값을 설정하는 것을 권장한다.

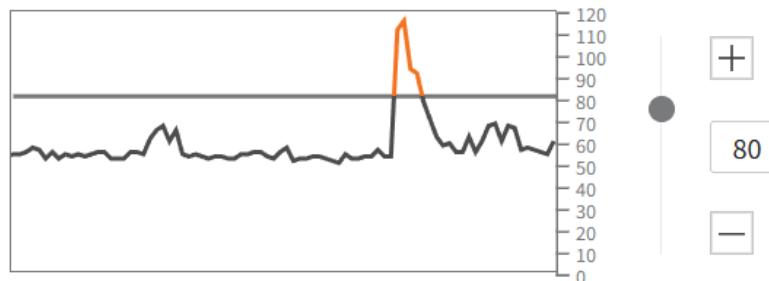


그림 6: 음원 분류 dB 임계값 설정 예시

그림 6 과 같이 사용자 설정에서 dB 임계값은 슬라이더나 숫자 입력 필드를 통해 직관적으로 조절할 수 있으며, 이 설정은 실시간으로 음원 감지 민감도에 직접적인 영향을 미친다. 그래프는 시간 경과에 따른 음원 dB 변화(검은색 선)와 설정된 임계값(Threshold, 회색 선)을 시각적으로 나타내어, 음원 레벨이 임계값을 초과할 때(주황색 피크) 이벤트가 발생함을 쉽게 이해할 수 있도록 돋는다.

4.4. 음원 방향 보정 및 시스템 설정

한화비전 제품은 이벤트를 오디오 클립으로 제공하며, 여기에는 음원 분류 및 방향 탐지 결과가 함께 표시된다.

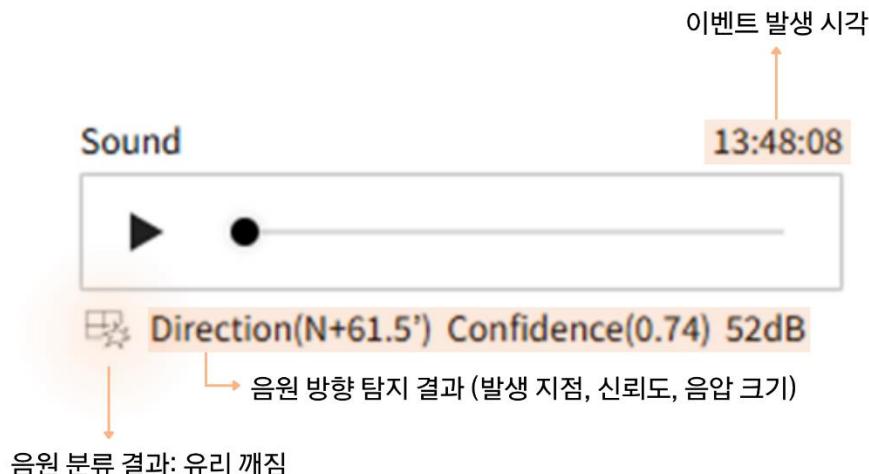


그림 7: 음원 방향 탐지 결과 예시

그림 7 과 같이 하단에 음원 분류 결과를 직관적으로 보여주는 아이콘과 함께 음원 방향 탐지 결과가 나타난다. 여기서 'Direction(N+301.8°)'는 음원 발생 지점의 방향이 북쪽(N)으로부터 301.8도 떨어진 곳에 위치한다는 것을 의미한다.

함께 제공되는 'Confidence(0.74)'는 탐지 결과가 74%의 신뢰도를 갖는다는 뜻이며, 음압 크기(52dB)와 함께 사용자가 상황을 정확하게 인지하고 신속하게 대응할 수 있도록 돋는다.

시스템이 제공하는 음원 방향 정보는 초기 설치 환경이나 시간이 지남에 따라 실제 북쪽 방향과 오차가 발생할 수 있다. 정확한 방향 정보는 시스템의 활용도를 높이는 데 필수적이므로, 필요에 따라 기준이 되는 N 값(북쪽)을 보정하는 것이 중요하다. 이를 위해 다음 3 가지 방법을 이용할 수 있다.

- 제품을 나침반 N 극이 가리키는 정북 방향으로 설치한다.
- 제품 내 [System] > [제품 정보] > [마운팅 모드] 메뉴에서 '북쪽과 장비 기준점 사이 각도' 필드에 나침반 N 극을 기준으로 카메라 기준점까지 시계 방향으로 측정된 각도를 직접 입력한다.
- Wisenet Installation 툴 내에 포함된 나침반 기능을 활용한다. 이 기능을 통해 보다 편리하고 정확하게 초기 방향 설정을 진행할 수 있다.

4.5. 복합 음향 환경 및 예외적 음원 활용 팁

- **복합 음향 환경에서의 활용:** 2 종류 이상의 음원이 동시에 발생하는 환경에서는 AI 모델이 하나의 음원으로 분류하거나 오분류할 수 있다. 이는 자연스러운 현상 이므로, 시스템이 제공하는 정보를 종합적으로 판단하면 정확한 상황 인지에 도움이 된다.
- **정확한 알람을 위한 환경 분석:** 음원 분류 모델은 쇠가 부딪히는 마찰음, 동물 울음소리, 악기 소리 등 이벤트 음원과 유사한 소리나, 분류 카테고리에 포함되지 않는 갑작스럽고 강력한 소리에 대해 알람을 발생시킬 수 있다. 이러한 모델의 특성을 이해하면, 예외적인 소리에도 발생할 수 있는 알람을 사전에 예측하고 대비해 불필요한 혼란을 효과적으로 줄일 수 있다.

5. 결론

한화비전 AI 오디오 솔루션은 단순히 보는 것을 넘어, 지능적인 소리 분석을 통해 기존 시스템이 감지하지 못했던 위협까지 포착하며 빈틈없는 조기 감지 체계를 완성한다. 음원 분류 및 방향 감지를 통해 빈틈없는 보안을 실현하며, 지속적으로 더 나은 솔루션을 제공하고 있다.

본 백서는 솔루션의 잠재력을 최대한 활용할 수 있는 실용적인 지침을 제공, 사용자가 각자의 환경에 최적화된 맞춤형 보안 시스템을 구축할 수 있도록 돋는다. 시스템 설치부터 성능 최적화에 이르기까지, 모든 단계에서 사용자가 불필요한 시행착오를 줄이고 즉시 최상의 결과를 얻을 수 있도록 명확한 가이드라인을 제시하고자 한다.

한화비전은 앞으로도 시장에 요구에 부응하고 사용자가 예측 불가능한 상황에서도 더욱 안정적이고 효율적인 보안 환경을 경험할 수 있도록 AI 오디오 분석 기술을 지속적으로 발전시켜 예측 불가능한 위협 대응을 강화해 나가고 있다.

Hanwha Vision Co.,Ltd
13488 **Hanwha Vision R&D Center**,
6 Pangyo-ro 319-gil, Bundang-gu, Seongnam-si, Gyeonggi-do
TEL 070.7247.8771-8 **FAX** 031.8018.3715
www.HanwhaVision.com