i-PRO AI Sound Classification Technology

November 2021

Panasonic i-PRO Sensing Solutions Co., Ltd.

i-PRO AI Sound Classification Technology

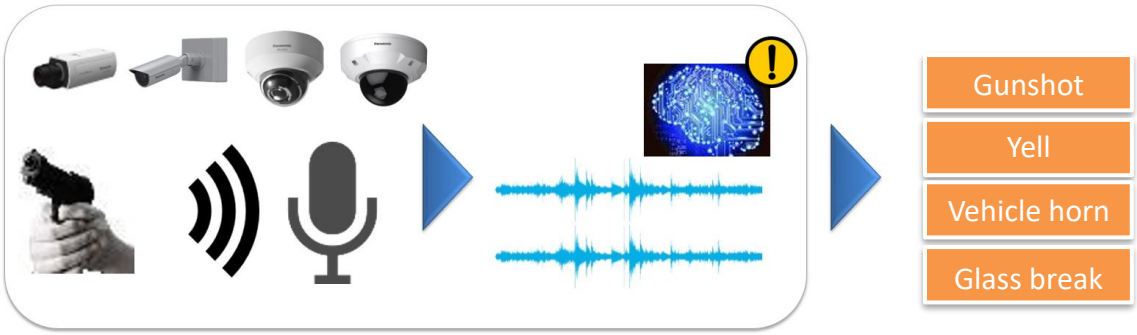
1. Introduction

The surveillance camera industry is promoting greater use of AI-based video analysis technology. With a surveillance camera and an AI video analysis app, suspicious individuals can be automatically detected and the individual's physical characteristics can be used to perform a video library search. This reduces the monitoring burden placed on the camera operator and improves efficiency.

Further surveillance efficiency can be realized by analyzing the camera's audio data together with the video data. Especially in a noisy environment, it is difficult for an operator to listen to sound captured by multiple cameras at the same time. Therefore, automatic audio analysis technology also offers a lot of promise.

One of the sound detection features of conventional surveillance cameras is a function that issues an alarm whenever the volume exceeds a specific level. In a noisy environment, this leads to many false alarms, putting considerable restrictions on audio surveillance at a practical level.

Just like with AI-based video analysis, by using AI sound classification technology to detect abnormal sounds such as people yelling, we can detect incidents earlier and search recorded video more effectively. This white paper provides an overview of i-PRO's AI sound classification technology using Deep Learning and explains settings to optimize the performance of AI sound classification.



Identifying specific sounds among other noise using Panasonic i-PRO's Deep Learning network

2. AI Sound Classification Technology Using Deep Learning

In the past, AI-based video analysis usually took the form of a server-based system that performed arithmetic processing on video uploaded to the server. In recent years however, Edge AI, which allows for Deep Learning arithmetic processing to be performed within each surveillance camera, is becoming widespread. The advantages of Edge AI include 1) cost reduction, 2) better information security, and 3) enhanced real-time performance. By installing various apps on the

i-PRO Network Camera, customers can easily adopt video analysis technology with Edge AI.

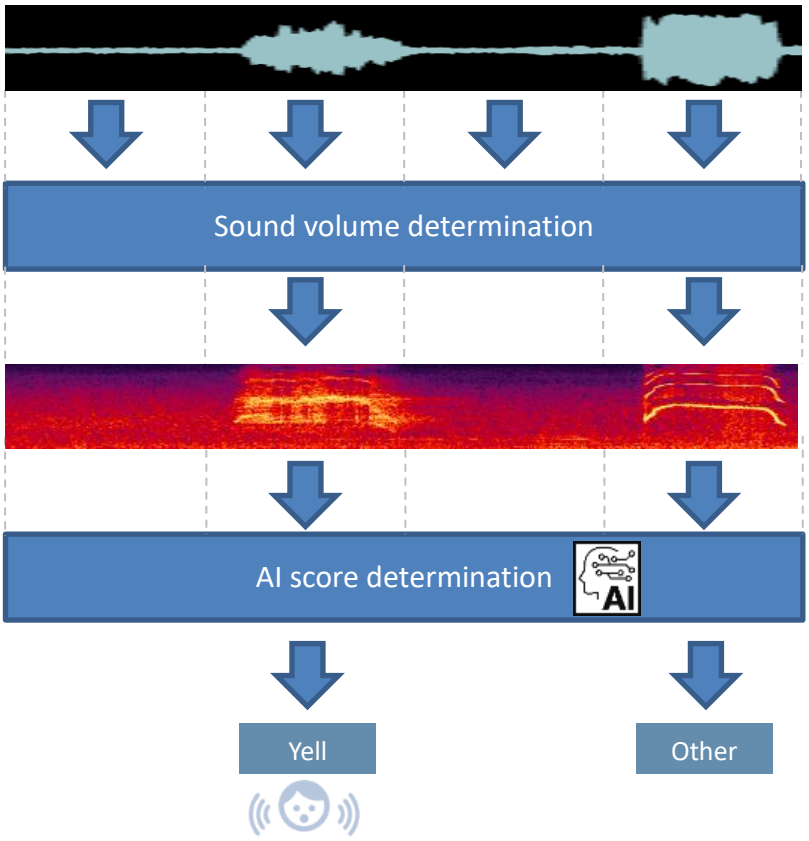
With the i-PRO Network Camera, both video and audio analyses are performed by Edge AI processing. Audio analysis generally involves a smaller amount of data than for video analysis and can be processed relatively easily. This means both video and audio analyses can be performed simultaneously within the camera.

The AI sound classification technology uses two indicators, the captured sound volume level and an AI score, to determine whether an alarm should be issued.

To identify a sound, the system first compares the captured sound volume level with a preset threshold value. If it is greater than the threshold, AI is then used to determine what kind of sound it could be. To come up with an AI score, the system determines whether the captured sound corresponds to any of four target sound categories: Yell, Glass break, Vehicle horn, and Gunshot.

To do this, the captured sound is divided into regular segments, signal processing is performed, and the feature quantity for the target sound is determined. By inputting the feature quantity into an AI model and performing inference calculation, the degree of similarity to any target sound is calculated. When the similarity level exceeds a certain value, target detection is output as an alarm.

As already mentioned, the AI model uses four types of target sounds: Yell, Glass break, Vehicle horn, and Gunshot. The model was created by providing the system with various sound sources and training it to learn the characteristics unique to each target sound. Moreover, by having the system learn to identify target sounds under various conditions, such as situations with typical environmental noise or other non-target sounds, the possibility of false positives caused by background noise is minimized.

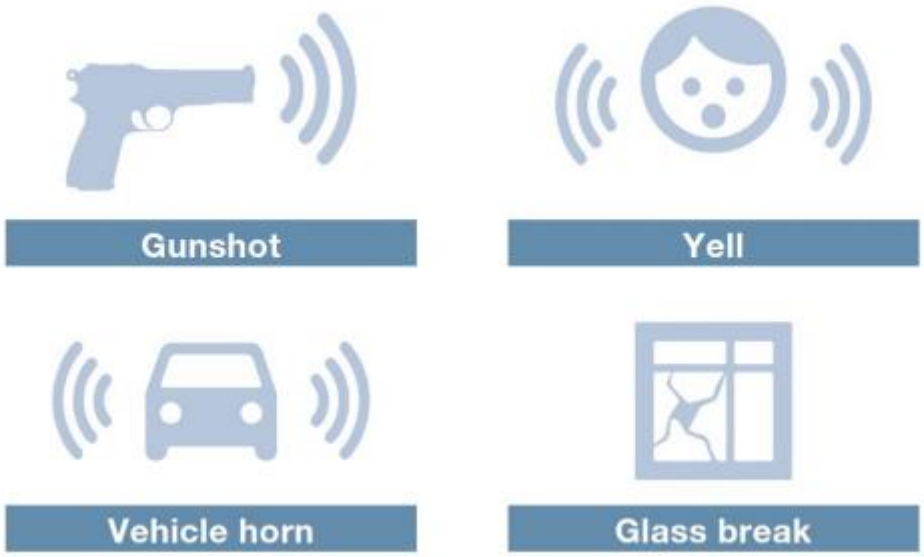


3. Installation Environment and Setting Procedure for AI Sound Classification

3.1. Detectable Sound Types

The four categories that can be selected as target sounds for detection are Yell, Glass break, Vehicle horn, and Gunshot. Whenever a sound is detected that matches a target category, an alarm is issued.

- Gunshot: Firearm discharge and reverberation sounds, as well as similar explosion/collision noises
- Yell: Female or male yells or screams, as well as children's screams, and various angry voices
- Vehicle horn: Honking that lasts for 1 second or longer
- Glass break: Sound of glass breaking and shattering



Multiple target sound categories can be selected at the same time. Since the alarm includes information on the sound category detected, the operator can use the Video Management Software (VMS) to search for video clips in which a specific target sound occurs

In addition, "Other" can be selected as the target sound category. When "Other" is selected, an alarm will be issued whenever the AI determination result does not correspond to any of the four categories above.

By selecting all categories (Yell, Glass break, Vehicle horn, Gunshot, and Other) as the target sound categories, any such sounds above a certain volume level will trigger an alarm. This setting is useful when any such loud sounds are regarded as abnormal.

3.2. Conditions of target sound for detection

The following can be set as conditions for AI sound classification detection.

1. The sound continues for 1 second or longer (excluding gunshots)
2. It is louder than the ambient noise level by 6 dB or more.
3. The sound is louder than a set volume threshold

With Condition 2 above, even if the ambient noise level is high, as long as the target sound is relatively louder, it can be detected by AI sound classification.

With Condition 3, detection sensitivity can be set using a volume threshold. When the sensitivity is set to "Low," only loud sounds are detected. In addition, unlike with Condition 2, the determination is based on the absolute volume. When a camera is installed indoors, it often records human conversation and many types of other noises. While the noise level can be high due to

echoing in the room, the target sounds for detection are also loud. Therefore, by setting the detection sensitivity to "Low" and targeting only comparatively loud sounds, better detection performance can be obtained.

3.3. Sound-detection distance

In general, sound intensity diminishes as it travels from the source to the camera's microphone. The further away the source, the more likely the sound is to be affected by various dampening factors such as obstacles, sound absorbency of the ground surfaces, and the weather.

The distance range for AI sound classification can be specified by requiring the target sound to be at least 6 dB higher than the ambient noise in the camera's environment.

As noted above, since there are many variables affecting the relationship between the distance and the amount of sound dampening, it is difficult to strictly define the sound-detection distance. However, a guide to sound-detection distances is provided below.

Ambient Noise Level	Camera Location	Sound-detection Distance (Estimate)	
		Yell / Glass break (90-110 dB)	Gunshot / Vehicle horn (110-130 dB)
70 dB	Noisy street, main road	4 m (13 feet)	15 m (49 feet)
60 dB	Ordinary conversation, quiet car	8 m (26 feet)	30 m (98 feet)
50 dB	Residential area in the daytime, office	16 m (52 feet)	Approx. 50 m (164 feet)(*1)
40 dB	Residential area at night, library	32 m (105 feet)	Approx. 50 m (164 feet)(*1)

(*1) Considerable distance variation possible depending on the surrounding environment, such as a building blocking the sound

3.4. External microphone requirements

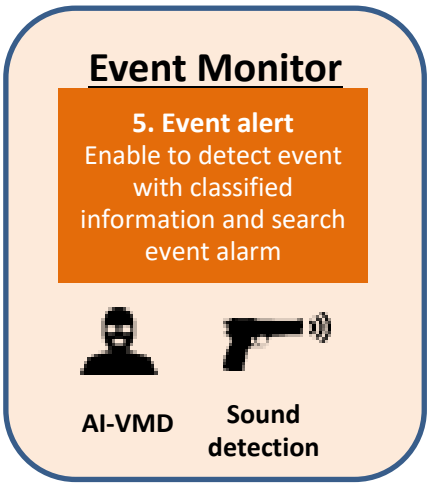
When using an external microphone for AI sound classification, a suitable type must be selected.

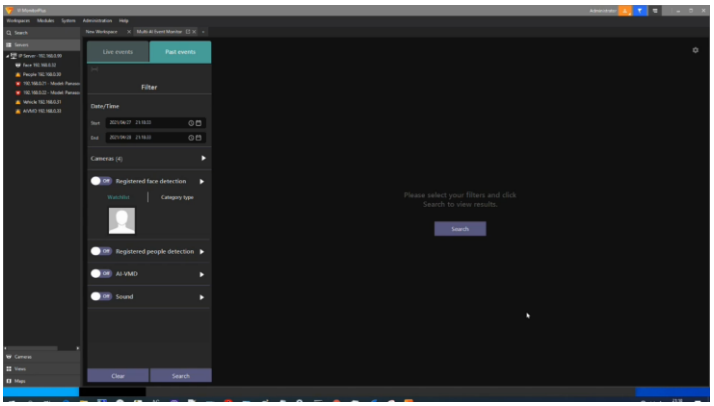
1. Frequency characteristics
   AI sound classification is in the range of 200 Hz to 8 kHz, and the frequency distribution of a captured sound is used as the feature quantity. Therefore, the microphone must be able to pick up frequencies in this range with a flat characteristic.
2. Directivity
   An omnidirectional microphone is required. If a microphone with strong directivity is used, the

accuracy of AI sound classification may decrease depending on the location of a sound source.

3.5. System linking

The VMS allows you to search recorded video using information from each camera alarm. Since an audio event caused by an abnormal sound is usually shorter than a video event such as intrusion by a suspicious person, a lot more time can be required for manual audio searches. By enabling video search using the sound alarm information, video surveillance costs can be greatly reduced.





4. Conclusion

By utilizing Deep Learning technology for audio analysis, detection can be limited to specific target sounds, even in noisy environments. The application of AI sound classification technology to surveillance cameras enables the creation of more efficient surveillance solutions. This includes VMS utilization of alarm information generated based on both video and audio detection processing.

About Panasonic i-PRO Sensing Solutions Co., Ltd.

Panasonic i-PRO Sensing Solutions Co., Ltd., is a global leader of advanced sensing technologies in the fields of Intelligent Surveillance, Public Safety, and Industrial/Medical Imaging. Established in 2019, i-PRO was built on a legacy of over 60 years of innovation with Panasonic.

The company's products, software and services extend human senses to capture moments of truth with innovations that inform and protect. In order to help create a safer world, Panasonic i-PRO Sensing Solutions Co., Ltd. supports the work of professionals who protect and save lives.

The Power of Truth

Panasonic i-PRO Sensing Solutions Co., Ltd.

4-1-62 Minoshima, Hakata-ku, Fukuoka-shi, Fukuoka 812-0017, Japan

ipro.panasonic.com