WISENET
SAMSUNG

White Paper:
Video/Audio Analysis Technology

Hanwha
Techwin America

hanwhasecurity.com

Contents

WISENET
SAMSUNG

Preface

To protect people and assets from accidents, incidents, and terror attacks, the number of video security cameras and systems being utilized is increasing at a high rate. However, the number of security personnel has not increased at the same rate, resulting in an average of 20 to 50 cameras being monitored by each security personnel. Viewing many monitors and cameras simultaneously prevents security personnel from focusing on core monitoring duties, which can lead to an increased probability of missing critical situations due to viewing fatigue.

For these reasons, interest in intelligent audio and video analysis technology for overcoming such limitations and for efficient monitoring is expanding, and research is striving to advance.

Intelligent audio and video analytics is a technology which alerts the operator of abnormal activities, detected through analyzing video and audio information, designed to prevent accidental or intentional actions and to minimize damage through prompt response.

Intelligent analytics can also be utilized during recording and search operations. Recorded events are tagged with the event type and associated metadata. An operator can simply search recorded video for specific event types to quickly locate an incident saving valuable time.

Hanwha Techwin is continuing to invest in the research and development of intelligent analysis technology, and this white paper is designed to provide information on the intelligent audio and video source analysis technology featured in Hanwha Techwin network cameras.

The following section details the various analysis techniques Hanwha Techwin provides.

2.1 Tampering Detection

Tampering detection is a technology which detects events that disturb normal monitoring, and it is a crucial technology which all monitoring systems must provide. In case of sudden changes, the camera may not be able to perform normal monitoring. If the following changes occur, check the device on site and implement suitable measures.

- Camera direction changed due to impact
- Camera focus significantly impaired
- Camera vision lost due to object being covered or the camera being spray painted on
- Camera video lost due to intentional blockage

In normal monitoring environments, small sudden or gradual lighting changes may be present, or the camera may be subject to repeated vibration due to wind or vibration from the installation location. Furthermore, an object temporarily appearing on the screen or repeated changes on portions of the screen may be detected.

Hanwha Techwin's tampering technology effectively excludes such elements in normal monitoring environments and is designed to detect only significant events. Furthermore, it provides real-time, visual screen change levels which allow optimization according to various monitoring environments for defining minimum detection time until sounding an alarm.

2.2 Face Detection

Face detection is a technology which identifies human faces from video images by identifying the key features of human faces. There are a variety of methods used to detect faces, including:

Template Matching Method
This method develops templates based on facial information extracted and registers the relationship in the system. Then it calculates the similarity between faces in video images and the templates.
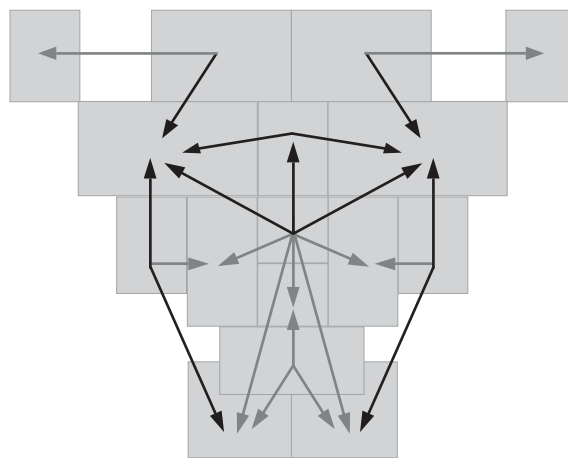

Figure 1. Sample template with facial and relation information

Feature Invariant Approach

This face detection method utilizes facial features which are less influenced by rotation, size and lighting changes. It combines information about eyes, noses and mouths to determine the presence of a person's face.
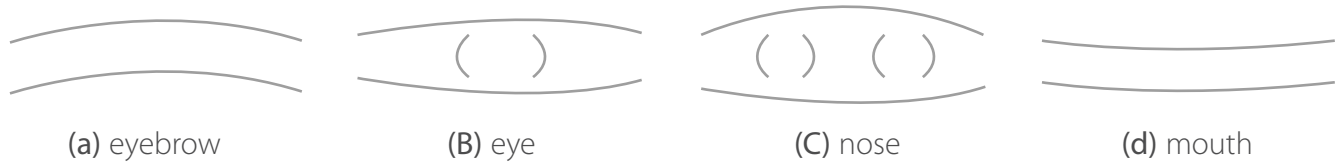

(a) eyebrow   (B) eye   (C) nose   (d) mouth

Figure 2. Facial feature information

Boosting Approach

This method utilizes basic patterns of faces which are compared to a classifier containing facial feature information for determining an individual's face.

Figure 3. Basic facial pattern sample

*Figure 1,2,3source: "Real-time face detection technology research trend" by Prof.Dongil Han of Sejong Univ.

Facial detection requires a significant amount of video information, and Hanwha Techwin's X Series vastly improves its face detection performance compared to previous products by collecting more detailed video from user designated areas. The Wisenet X series cameras require only 25x25 pixels to detect a face compared to 90x90 or 45x45 of previous generation of cameras. The improved detection requires 3.25 times fewer pixel for 2 megapixel detection and 13 times fewer pixels for 5 megapixel detection. Thus, face detection can work on wider scenes and cases where the subject is farther away from the camera. Furthermore, the face detection function can detect up to 35 faces at one time.

Table 1 - Comparison of Face Detection Performance in Wisenet X and Previous Generation Models

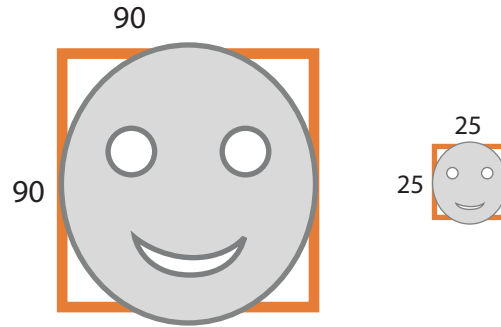| Input Video	Min. Detection Size		Ratio to Full Size Video
	Width	Height	
Previous 5M Camera	90	90	0.165%
Previous 2M Camera	45	45	0.098%
X Series 5M Camera	25	25	0.013%
X Series 2M Camera	25	25	0.030%

Figure 4. Face Detection Size of Previous Product (L) and X Series (R)

Accurate detection can be customized by configuring a detection area, exclusion areas, and sensitivity.
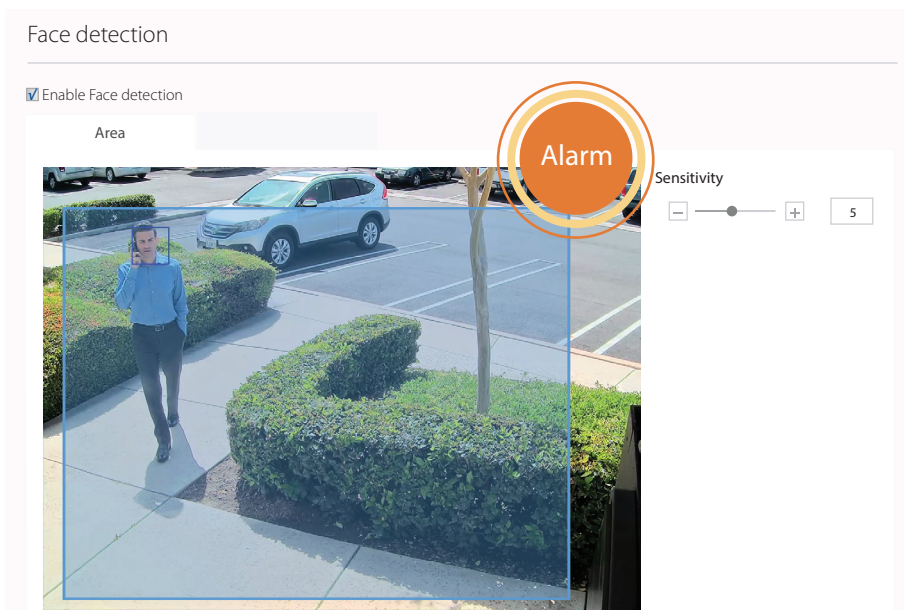

Figure 5. Face Configuration Detection Options

2.3 IVA (Intelligent Video Analysis)

The system can be set to generate an event and take an action in cases where movement is detected or a situation that satisfies the defined event rules.

Common settings include a user-defined minimum and maximum object size. To avoid detection errors due to noise and extraneous movements, set a suitable minimum/maximum detection size for the installation environment. However, as identical movement from identical locations may be detected differently, be sure to include margins in the minimum/ maximum size limitations. A sensitivity adjustment setting is available to change the threshold for movements to be detected. To detect events only in the specified areas, exclusion zones are available to ensure detection only occurs where desired and to prevent false positive events detections.

Sensitivity		— ●———— +	80	
Size	■ Minimum	174	X	174
	■ Maximum	1729	X	1729

Figure 6. IVA Common Configuration Options

Virtual Line Crossing Detection

Objects crossing a designated virtual line can be detected. The direction of detection can be specified. Configuration options include defining virtual lines and direction.

Crossing Line Detection

Figure 7. Virtual Line Crossing Detection

Line   `1` `2` 3 4 5 6 7 8

Categories   ☑ Crossing

Direction   ○ A ➡ B
○ A ⬅ B
◉ A ↔ B

Figure 8. Virtual Line Crossing Configuration Options

Enter / Exit Detection

Objects entering / exiting a designated virtual area can be detected. Configuration for all virtual area based IVA is performed in on a single configuration screen, shown in figure 9.

Enter / Exit Detection

Figure 9. Enter / Exit Detection Examples

Area   `1` 2 3 4 5 6 7 8

Categories   ☑ Intrusion

☐ Enter

☐ Exit

☐ Appear(Disappear)

Minimum duration (s)   — ●——— +   10

☑ Loitering

Minimum duration (s)   — ●——— +   10

Figure 10. IVA Virtual Area Based Configuration Option

Intrusion Detection

Intrusion detection can trigger an event when movement is detected within a designated virtual area.
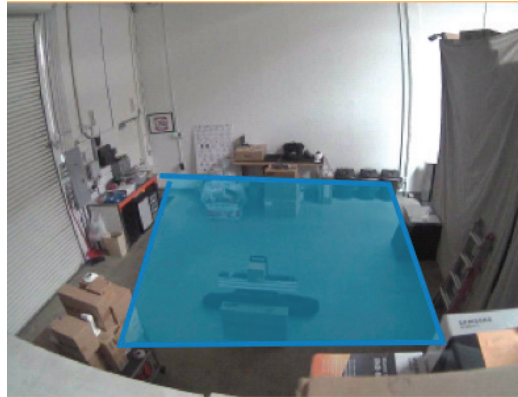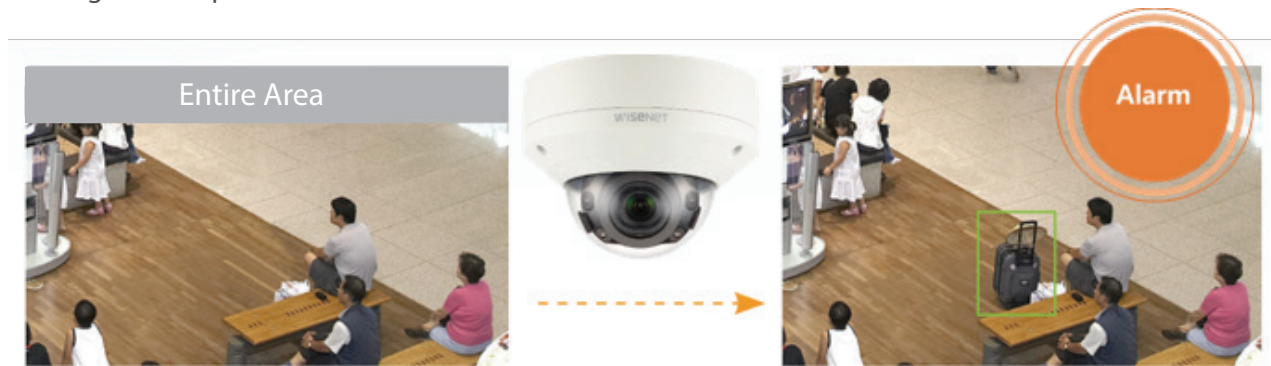

Figure 11. Intrusion Detection Example

Appearing Detection

Objects appearing in a designated virtual area and holding their position for more than the set observation time are detected. Configuration option includes detection area and minimum detection duration.

Appearing Detection

Figure 12. Appearing Detection Example

Disappearing Detection

Objects disappearing from a designated virtual area and remaining absent for more than the set observation time are detected. Configuration option includes detection area and minimum detection duration.

Disappearing Detection

Figure 13. Disappearing Detection Example

Loitering Detection

Objects loitering in a designated virtual area for more than the set observation time are detected. The camera looks for movements of similar patterns that are contained within the virtual area. Once these patterns are observed for a specified duration, then loitering detection is triggered. Configuration option includes detection area and minimum detection duration.
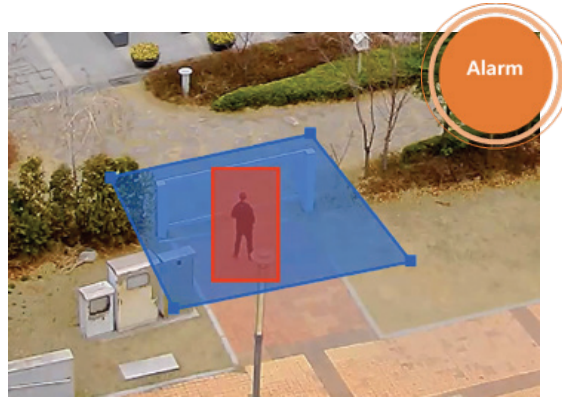

Figure 14. Loitering Detection Example

2.4 Audio Detection

Audio detection is a technology that detects audio levels which exceed the user-defined levels. As audio levels are greater in abnormal situations than in normal situations, audio levels exceeding set levels are detected as being an abnormal situation. Through audio detection technology, the camera is able to detect abnormal situations, then notify the operator via event signals allowing the operator to take suitable measures.

Hanwha Techwin's audio detection technology calculates the absolute level of actual audio signals collected using the microphone, then normalizes the levels in steps of 1 to 100. It defines the normalized level as the audio size, and audio levels exceeding the set level are detected as an event. Note that the audio size used for this purpose does not correlate to specific decibels (dB) values.
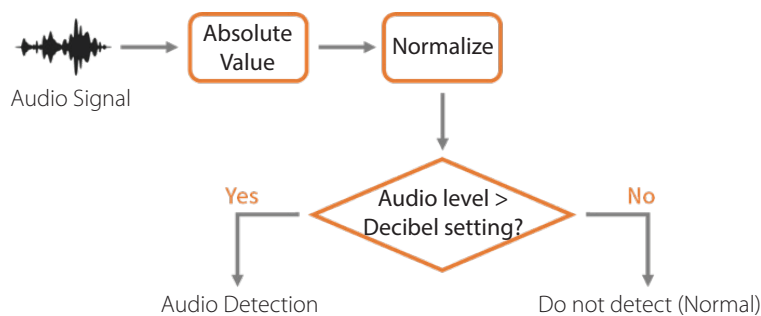

Figure 15. Audio Detection Algorithm Flow

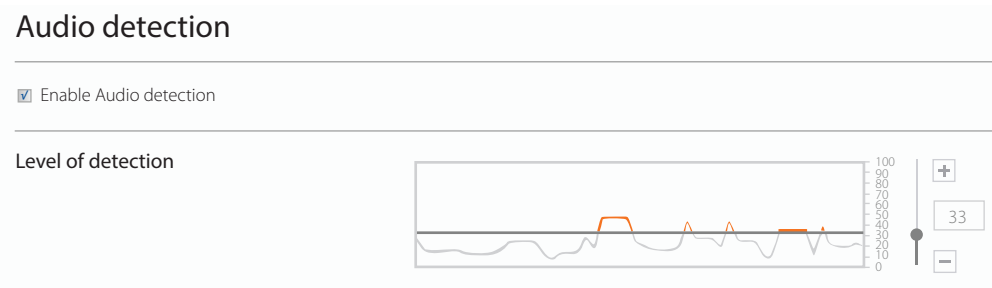Configuration involves adjusting the audio level of detection threshold as needed.

Audio detection

☑ Enable Audio detection

Level of detection

Figure 16. Audio Detection Configuration Options

2.5 Audio Source Classification

Audio source classification is a technology to classifying audio being input to the camera. Since the audio detection technology previously discussed generates alarms based simply on audio size, it may generate events even under normal situations. To overcome such limitations, technologies to classify audio source types have being developed.

When the camera classifies the audio source type satisfying the criteria defined by the operator, it then notifies the operator via event trigger allowing a suitable response to be taken.

Hanwha Techwin features an audio source database which supports the classification of screams, gunshots, explosions and crashing glass. The camera extracts the characteristics of the audio source collected using the camera's internal or externally connected microphone and calculates its likelihood based on the pre-defined database. It selects the audio source with the highest likelihood and generates an event. The algorithm flow classifies audio sources as follows:
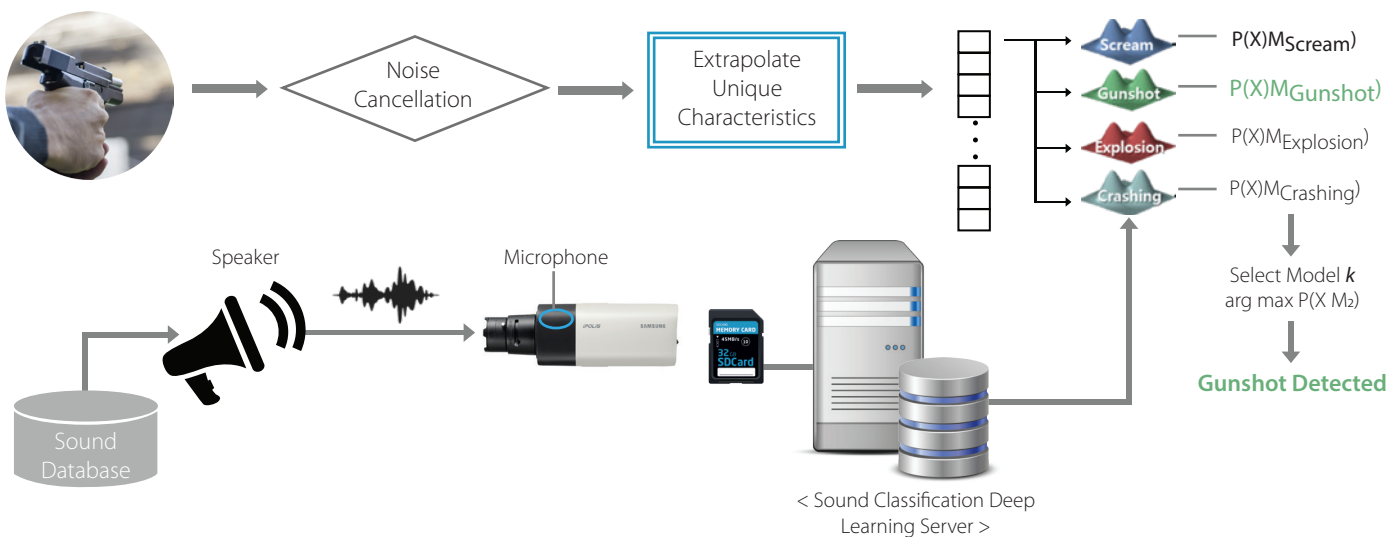

Figure 17. Audio Source Classification Algorithm Flow

Hanwha Techwin's audio source classification technology available in X Series cameras features three customizable settings for category, noise cancellation and detection level for optimum performance in a variety of installation environments. It also provides a graph which visualizes audio source levels to allow for the intuitive checking of noise cancellation and detection levels setup.

1) Category
Generates events based on audio source type detection. An operator can select the type of audio source for detection, and multiple audio sources can be selected.
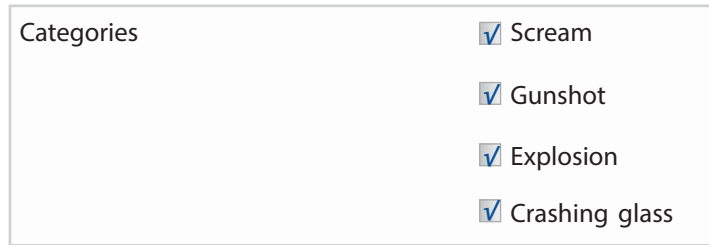

Figure 18. Audio Source Detection Category Setup

Scream: Generates events based on detections of loud voices such as screaming and yelling of adults and children.
Gunshot: Generates events based on the detection of non-continuous gunshot sounds.
Explosion: Generates events based on the detection of explosion sounds.
Crashing Glass: Generates events based on the detection of crashing glass sounds.

2) Noise Reduction
Depending on the environment where the microphone is installed, the operator can enable the Noise Reduction function. This function can reduce background noise greater than 55dB-65dB for increased detection accuracy.

Using the level of detection graph, the user can enable or disable the noise reduction function to view the result and validate the optimum configuration.

With noise reduction enabled, the system analyzes the attenuated audio source. As such, the audio source classification performance may be hindered or generate errors.
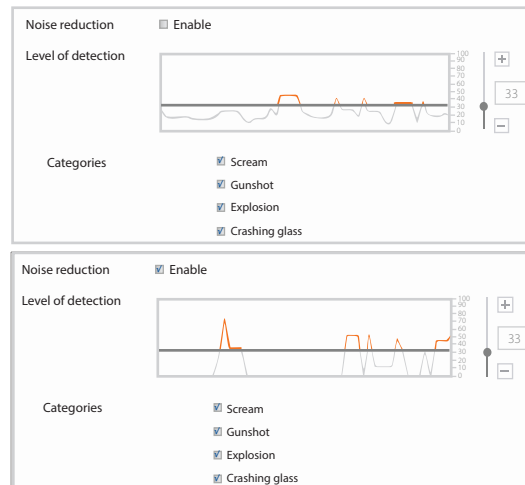

Figure 19. Audio Source Level Graph with Noise Reduction Disabled (top) and Enabled (bottom)

3) Detection Level

The detection level specifies the audio source volume levels at which to perform audio source classification. Audio volume levels of the audio source are updated continuously and displayed on a graph with the most recent indications on the right. Audio source classification is performed only on audio sources exceeding the set level. Thus, only input audio sources with volume levels exceeding the threshold undergo audio source classification.

Lower thresholds results in greater audio source classification data and possibly a greater misdetection probability. Higher thresholds results in less audio source classification data and greater non-detection probability. The threshold must be set appropriate to the surrounding noise level of the camera.
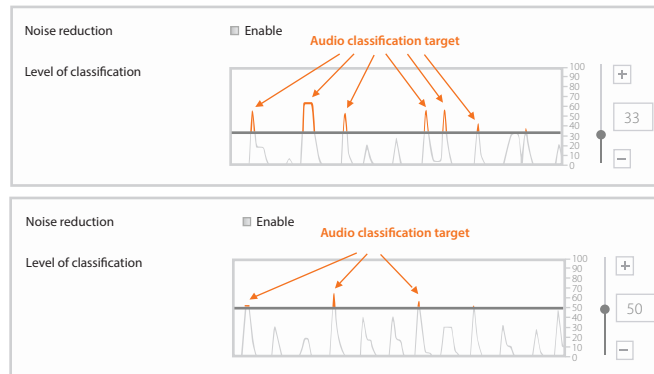
Figure 20. Audio Source Classification targets for Level 33 (top) and Level 50 (bottom)

2.5   Image Stabilization

Image stabilization is a technology that compensates image shaking due to vibrations from the environment to produce a stable image. In general, image stabilization technology is classified as a hardware method which utilizes the camera lens or image sensor to compensate for shaking, whereas DIS (Digital Image Stabilization) which utilizes a software analysis of shaking based on the image. As DIS compensates for shaking with software unlike hardware compensation methods, it can reduce a product's price by reducing the amount of hardware in the product.

As more cameras are installed in buildings, parking lots, roadways, and sporting areas, and as the technology for image analysis advances, image stabilization technology is being regarded as a crucial feature for enabling the various technologies contained to perform at their maximum capacity beyond simply recording stable images.

Hanwha Techwin's image stabilization is based on the software compensation method, DIS. The DIS featured in previous models of Hanwha Techwin is implemented in a 2-stage image compensation of vector estimation and movement compensation. It calculates the movement level between two continuous frames and compensates for shaking in two axes (Pan and Tilt). This technology does not require any additional hardware giving it an advantage in cost, but as it is highly dependent on the image, it has the disadvantage of malfunctioning due to movement vector estimation errors.

For this reason, Hanwha Techwin developed a gyroscope sensor integrated DIS technology to reduce malfunctioning and improve DIS accuracy. Independently operating gyroscope (gyro) sensors collect camera shake information aside from the movement vector information collected through image analysis reducing the probability of malfunctioning.
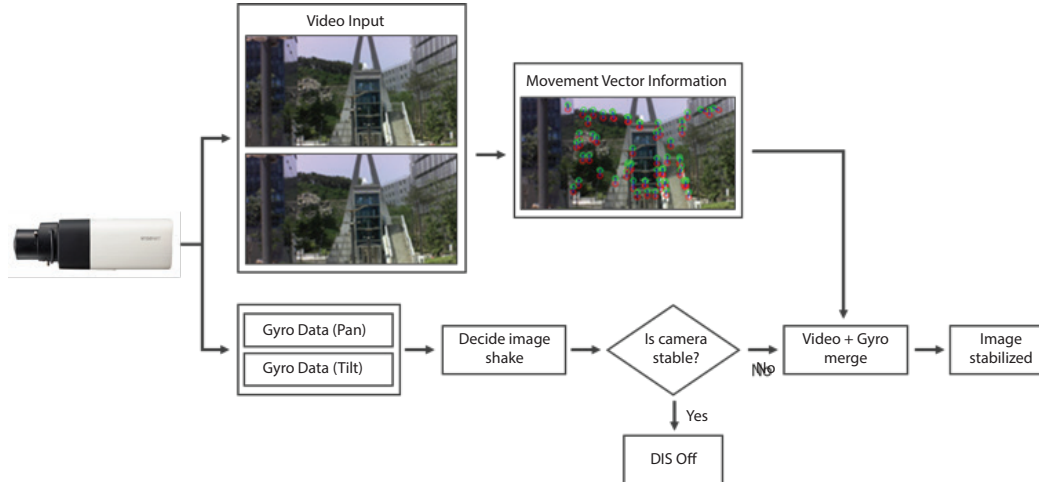
Figure 21. Gyro Integrated DIS Algorithm Flow

The gyro integrated DIS function is classified into two movements according to camera shaking.

1) Signals generated from the gyro sensor determines camera shake levels. If there is no shaking, the stabilization function is temporarily disabled. As a result, movement vector estimation with the camera stopped can reduce the probability of malfunctioning due to error.

2) During compensation, movement vector information of the gyro sensor is utilized to compensate movement vector estimation based on image analysis. The gyro sensor information is based on angular velocity. By modifying this angular velocity to pixel value, the movement vector between two continuous frames is estimated. DIS performance is dependent
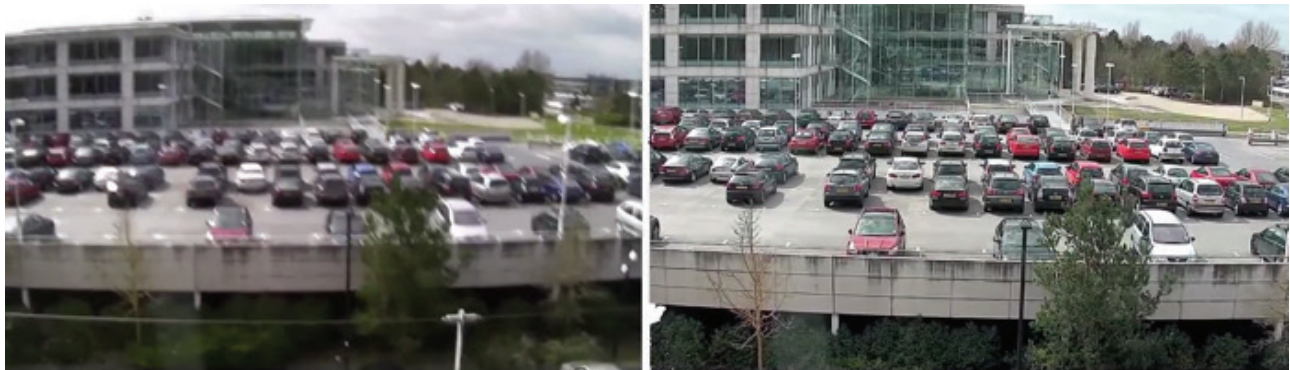

Figure 22. Digital Image Stabilization Example, DIS Off (left), DIS On (right)

Conclusion

The intelligent audio and video analysis technology featured in Hanwha Techwin's network cameras automatically notifies the operator of pre-defined situations detected. Through these technologies, they are able to not only monitor all cameras 24/7, but they can also ensure efficient operations by easily confirming and determining the circumstances of an event. With the use of intelligent audio and video analysis technology, a single individual can monitor many more cameras and monitors, as well as reducing the amount of missed critical events and enabling a prompt response. Furthermore, operators can review recorded video quickly by filtering or skipping to specific event types, easing the burden of reviewing all video or events, increasing operational efficiency.

W'ISENET
SAMSUNG

Hanwha
Techwin America

W'ISENET
SAMSUNG