Hanwha Vision

White paper

Behavior Detection

October 2026

Copyright © 2026 Hanwha Vision Co., Ltd. All rights reserved.

Contents

1. Introduction
2. Background
3. Technology
   3.1 Human Posture and Body Shape Analysis
   3.2 Object Tracking and Time Series Analysis
   3.3 Real-world Speed Calculation Based on Camera Calibration
4. Behavior Detection Functions
   4.1 Fighting Detection
   4.2 Hands Up Detection
   4.3 Running Detection
5. Installation Guide
   5.1 Recommended Camera Installation Conditions
   5.2 Recommended Installation Environment
   5.3 Region of Interest (ROI) Setting and Calibration
6. Conclusion

1. Introduction

Behavior detection, emerging as a core technology in modern video surveillance and monitoring systems, goes beyond single-image analysis to recognize complex human behaviors in real time by analyzing continuous video frame sequences. Unlike traditional reactive security systems, **Behavior detection** enables proactive security measures by identifying anomalous events such as **Fighting, Hands up, and Running** thereby driving a paradigm shift toward preventive security frameworks.

To realize high-performance **behavior detection** on edge cameras with limited embedded computing environments, Hanwha Vision has innovatively combined three core technologies: **Human Posture and Body Shape Analysis**, Multi **Object Tracking and Time Series Analysis**, and **Real-world Speed Calculation Based on Camera Calibration**.

In particular, by leveraging an Edge AI-based camera analytics approach, the system achieves real-time accuracy while fundamentally resolving high operational costs and network bandwidth constraints associated with large-scale server infrastructures.

This white paper introduces the technical leadership and differentiated behavior detection capabilities of Hanwha Vision's AI cameras. It presents the detection principles for each behavior, application scenarios, and an optimal installation and calibration guide to help achieve maximum practical performance in real-world security environments.

2. Background

AI-driven video analytics is rapidly evolving beyond static object detection and classification. Consequently, the paradigm of the video surveillance market is shifting from post-incident forensic investigation to real-time, **proactive accident prevention**.

Early Video Motion Detection (VMD) relied on simple pixel changes, resulting in frequent false alarms. While early AI-based **Object detection** and **Virtual line/area** analytics were introduced to address this limitation, they remain restricted to analyzing single frames to verify the presence of objects or basic motion. Modern physical security demands more than merely confirming human presence; it requires understanding what action a person is performing. Accurately recognizing human behavior requires analyzing continuous motion and situational context over time, rather than relying on isolated snapshots. Therefore, preemptively preventing incidents by identifying specific actions and posture transitions in real time has become the foremost challenge in the video surveillance industry.

'Behavior' cannot be determined merely through simple object displacement within a virtual zone on a single frame. It fundamentally requires the analysis of sequential video frames over time to estimate posture and understand temporal scene dynamics. However, centralizing these computationally intensive workloads on large-scale AI servers introduces severe architectural bottlenecks, including massive network bandwidth demands and high operational costs. To substantially reduce data transmission overhead and enable deterministic, low-latency response, deploying an Edge AI architecture is essential.

The critical hurdle lies in overcoming hardware constraints. Because on-camera embedded environments operate under strictly constrained CPU and NPU resources compared to centralized servers, lightweight deep learning model design and inference optimization—enabling practical, high-throughput temporal frame analysis directly on the device—have emerged as core technological differentiators.

3. Technology

Hanwha Vision's AI cameras leverage the powerful built-in computational capabilities of the NPU (Neural Processing Unit) to analyze complex human motions and interactions in real time directly at the camera edge. Because conventional single-frame analysis can determine only the presence or absence of objects and cannot accurately identify specific incidents, it has become necessary to understand and interpret scene changes over time through multi-frame analysis. The core algorithmic mechanisms used to precisely detect object behavior include human pose and body shape analysis, time-series analysis based on object tracking, and velocity estimation.

3.1 Human Posture and Body Shape Analysis

To overcome the limitations of recognizing human position merely as a rectangular bounding box, **Human Posture Analysis** technology is applied to precisely model the overall shape and morphological features of the human body. This technology uses deep learning networks to infer overall body shape transformations from 2D image data, including upper/lower body proportions as well as extension and flexion of body extremities, and reconstructs this into a dynamic pose model in real-time.

Based on the constructed human shape data, biomechanics are analyzed for each action. For example, **Hands up** is distinguished from everyday actions by measuring the extent and duration that the arm area clearly extends upward along the vertical axis (Y-axis) beyond a reference line based on the body's central axis and upper portion. **Fighting** identifies violent behavior by analyzing complex interaction data in situations where multiple objects are positioned in close physical proximity, including rapid and repetitive extensions of body outer contours, irregular oscillation of center of gravity between individuals (jittering), and sudden posture collapses.

3.2 Object Tracking and Time Series Analysis

A single frame's static posture alone cannot define the complete context of an action; therefore, continuous tracking of posture changes over time is necessary.

First, **Multi-Object Tracking (MOT)** assigns unique IDs to multiple people within the scene and maintains uninterrupted trajectories as frames progress. The extracted human pose data is accumulated as time series data in 'time window' units. A deep learning-based temporal classifier then analyzes frame-by-frame body deformations and movement speeds to make a final determination on whether the current motion matches a predefined action pattern.

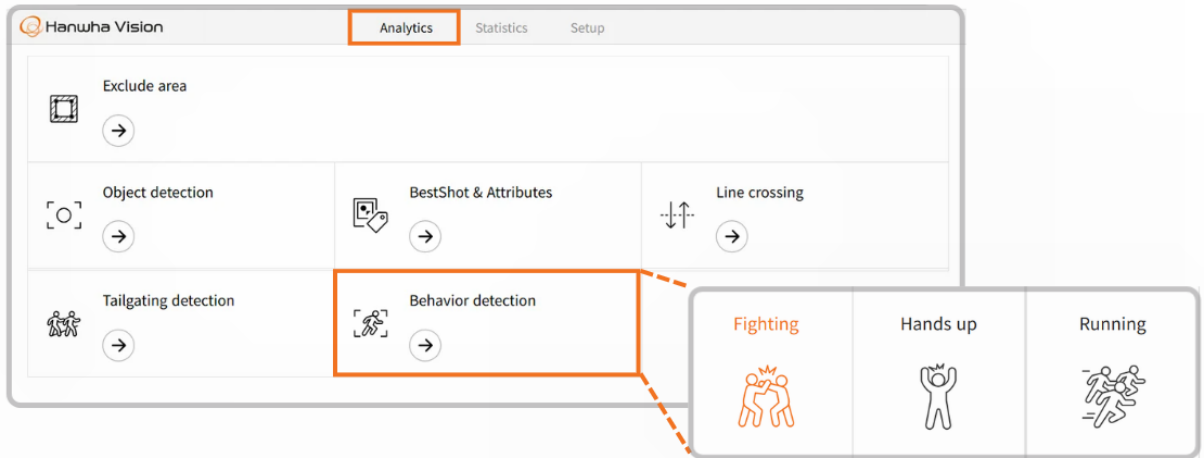
3.3 Real-world Speed Calculation Based on Camera Calibration

Running detection requires accurate measurement of not just posture changes but also the actual physical movement speed of the object to effectively prevent false positives. Generally, perspective distortion occurs: objects close to the camera show large pixel movement even with minimal movement, while distant objects show small pixel movement even when moving rapidly.

To correct this, 2D screen coordinates are converted to 3D real-world ground plane coordinates based on the camera's physical installation height, tilt angle, and focal length using perspective transformation. Through the generated perspective transformation matrix, frame-by-frame pixel movement distance of the lowest point of the human body shape (foot contact with ground) is calculated in real-time as actual physical velocity (m/s). Running is precisely identified by evaluating whether the human's actual movement speed continuously exceeds a preset threshold.

4. Behavior Detection Functions

To utilize behavior detection, verify that **Behavior detection** is specified under the Analytics section of the camera's datasheet. **Behavior detection** requires enabling the **BestShot** feature within the on camera **WiseAI** application, with **Person** designated as a mandatory detection target.



While posture detection runs continuously on incoming frames, individual or group behavioral detection models are called alternately according to defined trigger rules and resource availability, balancing the limited CPU/NPU resources in embedded environments with detection performance.

For major behaviors such as **Fighting**, **Hands up**, and **Running**, it is recommended that the object size be at least 15% of the total frame to clearly identify the behavior. Performance accuracy increases as the background becomes simpler and lens distortion decreases. Installation environment should be verified, and when special environments such as low light, backlighting, or focus changes are chronic issues, repositioning the installation location should be considered first.

4.1 Fighting Detection

The **Fighting** detection feature is a technology that recognizes physical collisions or violent interactions between two or more individuals in real time. Unlike simple crowd congestion or routine conversation scenarios, it captures irregular and rapid human body shape deformations and abnormal inter-object contacts that accompany aggressive behavior, enabling early warnings for security threats. As a result, high utility is expected for blind-spot monitoring in public spaces, nightlife areas, and schools.

Furthermore, in specialized high-security facilities where physical conflict can lead to critical consequences—such as prisons, detention centers, and psychiatric hospitals—this technology plays a crucial role in minimizing damage by being the first to identify violent incidents and enabling the immediate deployment of security and rescue personnel.

To detect fighting, at least two subjects must be present within the camera's field of view (FoV) in close physical proximity. An event is definitively classified as a fight when dynamic behavioral patterns—such as strikes (punching/kicking), swinging weapons, assaults on a downed individual, rapid limb extensions, abrupt center-of-gravity shifts, and postural collapse—persist for a defined duration (5 seconds). To accurately analyze human body biomechanics, the upper and lower body movements of the target individuals must be exposed in the video at a sufficient size.

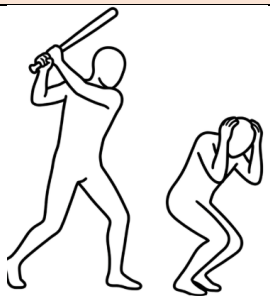
Detectable Postures			**Undetectable**
			
Punching/kicking	Swinging weapons	Assaulting a downed individual	Hugging, or tangled to a degree of wrestling

Table 1. Fighting and Assault Detection Status by Posture

If objects are severely occluded or in close physical contact at wrestling levels, it becomes difficult to clearly identify individual objects, and fighting may not be detected. Additionally, false positives may occur from body movements that are difficult to distinguish from fighting (dynamic dancing, rough play, etc.).

For more accurate fighting and assault detection, in environments with extreme illumination reduction or where subjects are too small to identify individual human body shapes in crowd congestion, analysis accuracy may be reduced. In such environments, installation location should be carefully considered. When multiple crowds of three or more are densely clustered and overlapping, individual body shape separation becomes difficult. In high crowd density situations, it is recommended to increase the number of cameras or distribute the field of view.

4.2 Hands Up Detection

The **Hands up** detection function is a technology that identifies specific actions of raising one or both hands above the head in response to external threats (surrender), exposure to coercive situations, or requests for emergency rescue. By serving as a visual emergency bell, it can recognize quiet rescue requests in situations where audio screaming is not detected. This is particularly effective in identifying urgent situations of shop owners and customers in high-risk robbery establishments such as convenience stores, banks, and jewelry stores, as well as late-night unmanned retail facilities.

Beyond criminal situations, it can also be utilized for industrial safety monitoring in large construction sites and ports, where workers manually signal toward the camera when equipment malfunctions or safety accidents occur, requesting immediate action from the monitoring center.

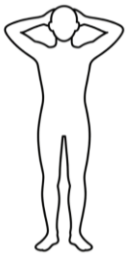
Detectable Hands Up Posture		Undetectable Hands Up Posture	
			
Posture with both arms fully extended above the head	Posture with hands placed on the back of the head	Posture with hands extended low	Posture with hands raised above the head but in motion

Table 2. Hands Up Detection by Posture

To detect **Hands up**, the camera continuously tracks the shape where the outer body region corresponding to the arm distinctly extends in the vertical (Y-axis) direction above the shoulder and head level, relative to the body's central axis. To prevent confusing this morphological change with daily movements such as stretching or reaching for objects in high places, an alarm is triggered only when the extended posture is stably maintained for a configured threshold time (5 seconds or more). A key feature of this temporal configuration is its ability to clearly distinguish momentary hand-raising actions—like greeting, stretching, or patting one's head—from sustained actions requiring a specific duration, such as surrendering or signaling for help.

	Preconditions for Hands Up Detection
Joint Visibility	Face, upper body, and both arms must be visible without obstruction.
Elbow Height	Both elbows must be positioned higher than the shoulders.
Posture Maintenance	The posture must be maintained for 5 seconds or more.

Table 3. Detailed Prerequisites for Hand-Raising Detection

4.3 Running Detection

Running detection identifies subjects moving at abnormally high velocities exceeding typical walking speeds, enabling the preemptive recognition of urgent flight, pursuits, or abrupt crowd surges. By combining dynamic human posture analysis with camera-calibration-based physical speed measurement, the algorithm effectively minimizes false alarms.

This feature serves as a valuable early-warning tool to prevent stampedes and mass panic triggered by terror threats or fires in high-density transportation hubs and commercial centers—such as airports, train stations, and large shopping malls. It is equally instrumental in tracking intruders rapidly fleeing sterile security zones, or enforcing pedestrian safety protocols in accident-prone areas where running poses direct hazards, including school corridors and escalator landings.

The running detection algorithm classifies an event as running when dynamic biomechanical transformations—such as stride elongation and forward torso tilt—coincide with a calibrated ground-plane physical velocity exceeding a predefined threshold (2.7 m/s) continuously for at least 2 seconds.

Because perspective foreshortening is transformed into real-world 3D coordinates to calculate speed, the contact area where the subject's feet touch the ground plane should ideally be clearly visible. However, even if environmental structures cause lower-body or foot occlusions, the algorithm precisely tracks the head apex and back-calculates both full-body anatomical proportions and ground contact coordinates, maintaining robust detection performance.

Accurate running detection requires precise knowledge of camera mounting height and tilt angle, which must be configured beforehand via calibration. When a subject moves directly along the optical axis (towards or away from the camera in the Y-axis), minimal lateral pixel displacement can degrade detection accuracy. To ensure dependable 3D ground-plane calibration, cameras must be installed with an adequate downward tilt angle, and Lens Distortion Correction (LDC) must be enabled to mitigate peripheral optical distortion.

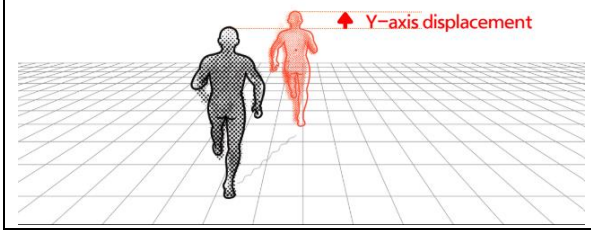
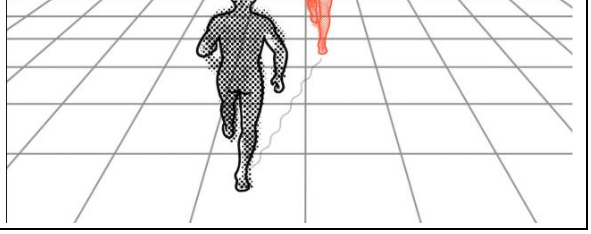
Small Tilt Angle	Large Tilt Angle
	

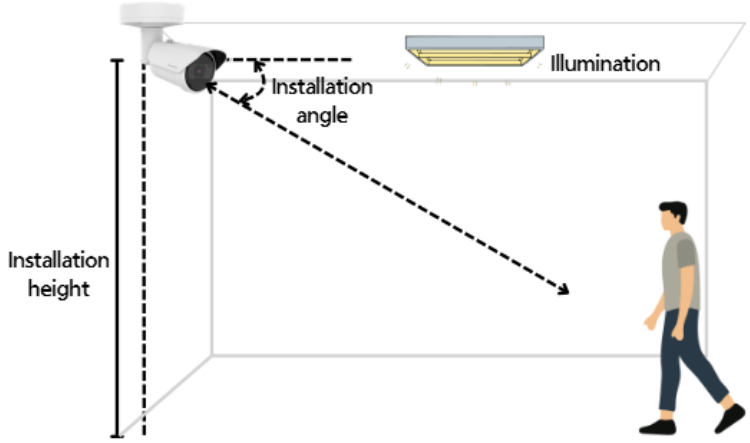
Table 4. Difference in Lateral Displacement Variation Based on Tilt Angle When the Subject Moves Straight Along the Camera's Optical Axis

5. Installation Guide

Behavior detection functions deliver optimal performance when installed and operated according to this guide. Because camera-embedded AI video analytics can be affected by ambient lighting and lens field of view (FOV), review the following recommendations to ensure maximum detection accuracy.

5.1 Recommended Camera Installation Conditions

Recommended installation conditions for stable **Behavior detection** are summarized in ***Table 5***. Avoid camera angles where subjects are frequently obscured (occlusion) by structures, large fixtures, or dense crowds. Since optical lens distortion is most pronounced at the edges of the frame, center the FOV to keep primary detection zones away from extreme peripheral areas.



Camera Installation Parameter	Key Installation Requirement
Installation Height	2.7 m (8.86 ft) or higher
Installation Tilt Angle	25°~45°
Roll Angle	0°
Illumination level	≥ 100 lx
Minimum Object Size	15% or more of the image vertical height recommended

Table 5: Recommended AI Camera Installation Conditions

5.2 Recommended Installation Environment

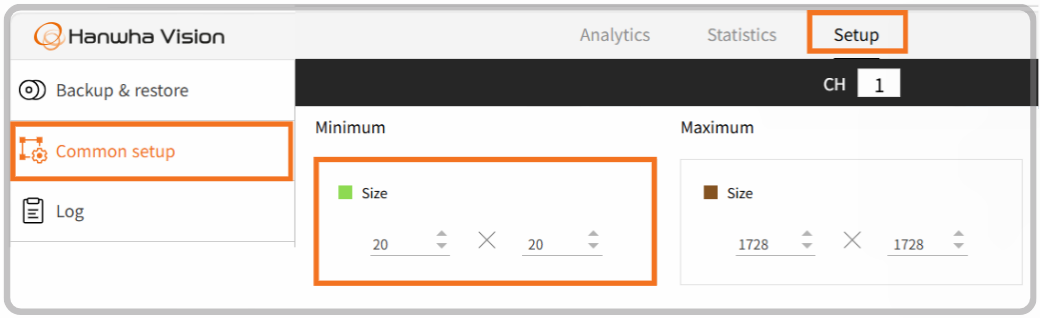
Stable behavior detection requires an environment free from obstacles that block body features. Because action recognition relies on dynamic limb movements, ensure arms and legs remain visible and that obstructions like foliage or pillars do not continuously conceal more than half of the human body.

Subjects must also clearly contrast with the background for reliable motion analysis. Strong tonal contrast between the target object and background is critical. Dense fog, direct sunlight on the lens (glare), or complex background textures matching the subject will hinder body segmentation and degrade performance. In areas with persistent low light, severe backlighting, or focus blur, prioritize improving scene illumination or repositioning the camera.

5.3 Behavior Detection Optimization Setup and Calibration

Configuring the target detection area as a Region of Interest (ROI) helps prevent false alarms by excluding unnecessary background areas. In particular, for **Hands up** detection, optical distortion at the screen periphery can cause arm positions or angles to be misrecognized; therefore, the ROI should be centered in areas where human body distortion is minimized.

Among the behavioral analytics functions, dynamic actions such as **Fighting** and **Running** are analyzed based on head position data. Accordingly, set the 'Minimum' object size to the lowest value under [Settings] > [Common Settings].



For Running detection, to measure actual physical speed, the camera's actual physical height and tilt angle must be accurately entered into the calibration settings, and the Lens Distortion Correction feature (LDC Fill Mode) must be enabled.

Because calibration assumes a flat horizontal surface, severe speed measurement errors will occur if the specified ROI includes ramps (slopes), stairs, step differences, or uneven terrain (such as mounds); the area must be strictly limited to flat ground surfaces.

Once initial setup is complete, always deploy personnel on-site to demonstrate the behaviors, verify detection rates and false alarm suppression, and finalize fine-tuning.

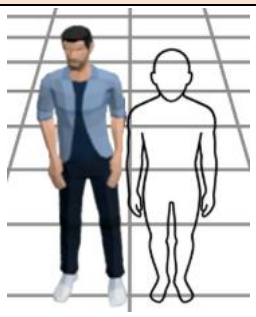
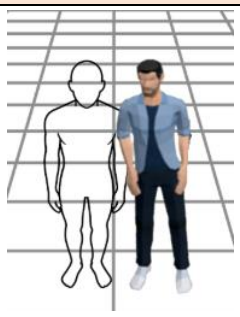
Measurement Results	Calibration Results
 Height: 3.1m, Tilt: 25°	 Height: 3.2m, Tilt: 25°

Table 6: Results of Calibration Correction

6. Conclusion

The modern video surveillance landscape has evolved from passive, post-event recording reviews into highly advanced, AI-driven proactive prevention systems. The **Behavior detection** technology discussed in this white paper transcends the mere detection of an object's presence or fragmented movements. By comprehending the dynamically changing context of complex human behaviors in real time, it elevates security systems to the next level.

Notably, Hanwha Vision has achieved innovative analytics capabilities even within the constrained computing resources of embedded camera environments (Edge AI). This was accomplished by successfully converging sophisticated algorithms: human pose analysis, time-series tracking, and real-world speed calculation based on camera calibration. As a result, it overcomes the budgetary and network bandwidth limitations typically associated with deploying high-cost, large-scale server infrastructure. Simultaneously, it rapidly and accurately captures only the critical anomalies that require immediate intervention—such as **Fighting, Hands up**, and **Running**. When installed according to optimal guidelines, Hanwha Vision's AI **Behavior detection** technology secures a critical window for immediate emergency response, serving as a vital tool to protect lives and property.

Leveraging this Edge computing expertise and advanced video analytics algorithms, Hanwha Vision delivers the most reliable real-time situational awareness solutions across diverse industrial and public security sectors. Moving forward, the company will continue to define the standards for the next-generation intelligent video surveillance market.

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 **FAX** 031.8018.3715
www.HanwhaVision.com

Copyright © 2026 Hanwha Vision Co., Ltd. All rights reserved.