"FlyingView®" -Real-Time Remote Monitoring System with Bird's-Eye View and AI Image Recognition-

Kazuki Kobayashi　Masashi Yuzue

From the perspective of addressing labor shortages and preventing infectious diseases, efforts toward unmanned on-site operation are accelerating with increasing implementation of automated and remotely operated vehicles, ships, robots, etc. In order to autonomously or remotely operate vehicles, an advanced remote operations system is necessary to allow an operator at a remote location to quickly make appropriate decisions and directly solve problems when an accident or problem occurs. Against this backdrop, there is an increasing need to see areas surrounding a vehicle with images that are easily viewable by humans from a remote location. Additionally, there is high expectation for AI to provide appropriate support in the operator's decision making and reduce/assist operational work.

Positioned as an AI-edge strategic product utilizing images and AI technologies, OKI's new "FlyingView® *1)" was developed to meet such demands. This article describes the features, functions, and configuration of FlyingView.

Features

FlyingView consists of a main unit to which four high-definition (HD) cameras are connected and a console connected via a network (**Figure 1**). As features of FlyingView, the intuitive "human eyes" and the comprehensive "AI eyes" will be explained.

(1) Intuitive "human eyes"

FlyingView combines high-definition (HD or FHD) images taken by cameras mounted on the front, rear, left, and right of a vehicle to generate bird's-eye view images and provides images seen from a virtual viewpoint (**Figure 2**). The virtual viewpoint images are distributed from the main unit to the remote console in real-time. Furthermore, using the console, the virtual viewpoint can be freely moved 360° around or above the vehicle (free viewpoint) enabling a bird's-eye view of the vehicle's surroundings from the height and direction the operator desires. This makes it possible to intuitively assess situations with the "human eyes" from a remote location.

(2) Comprehensive "AI eyes"

In FlyingView, an AI edge platform notably represented by OKI's AE2100[1] is utilized. Through the installation of an AI engine/application developed according to the detection target and operating environment as a container application, AI can constantly monitor camera images covering a full 360° of the surrounding area. As a result, the "AI eyes" can comprehensively capture happenings in places that are often overlooked by human eyes or scenes that pass by in an instant. Extraction and notification of scenes that need to be confirmed by human eyes will prevent human oversight.

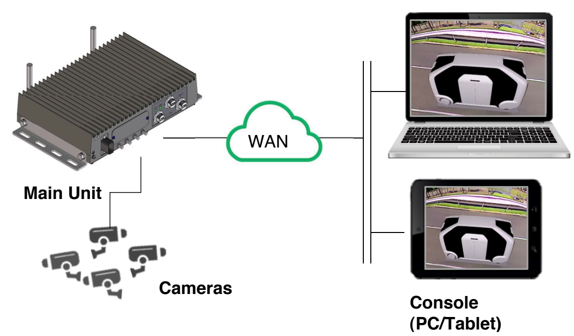

Figure 1. FlyingView System Configuration

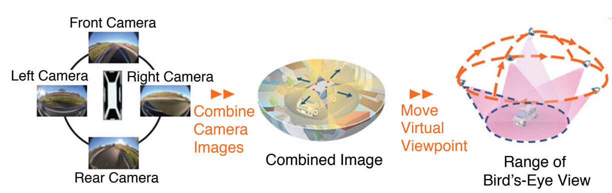

Figure 2. Virtual Viewpoint Image Generation

Functions

(1) Live Function

This function allows the remote operator to move the virtual viewpoint and see the situation around the vehicle in real-time from a free bird's-eye viewpoint (**Figure 3**). The operator can swipe his/her finger across the console screen or drag the mouse to intuitively move the viewpoint

*1) FlyingView is a registered trademark of OKI Electric Industry Co., Ltd.

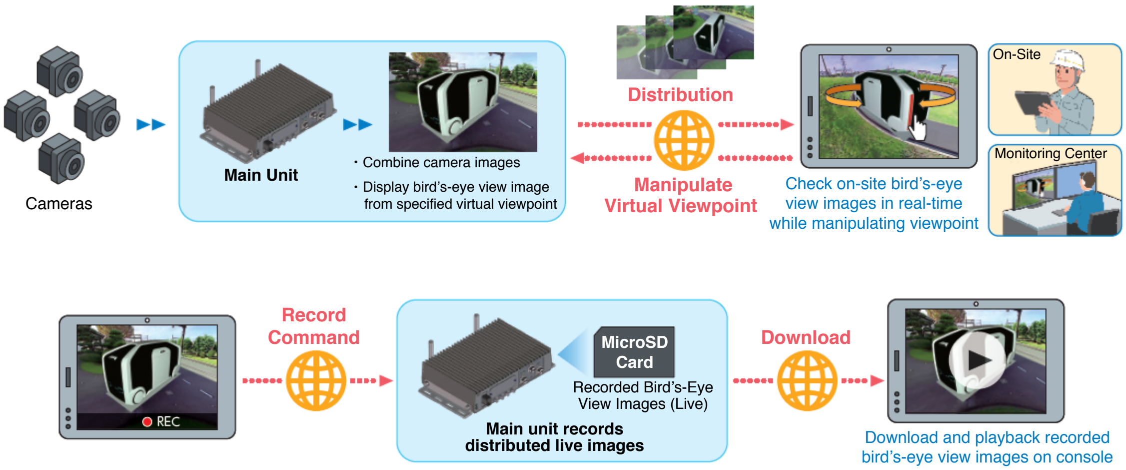

Figure 3. Live Function

and see places or direction he/she desires. This realizes the intuitive "human eyes." Additionally, the images that the live function distributes can be saved as a video file in the main unit.

(2) Playback Function

This function allows the remote operator to view prior situations surrounding the vehicle from a free bird's-eye viewpoint using the console in a similar manner as the "live function." The main unit is capable of constantly recording the front, rear, left, and right images from the cameras, and if necessary, generates bird's-eye view images from the recorded images and distributes the images to the console (**Figure 4**). Using this function, it is possible to intuitively grasp with the "human eyes" not only the real-time situation but also the situation immediately before an accident or problem occurrence.

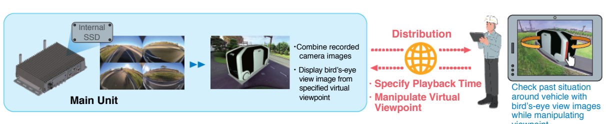

Figure 4. Playback Function

(3) Display Mode

The images displayed on console screen can be selected from one of the three available display modes. "Standard view" provides a full-screen display of a single bird's-eye view image. "Dual view" splits the screen to display two bird's-eye view images. "Multi-view" divides the screen into three parts to display one bird's-eye view image and two camera images.

The bird's-eye view of the dual view's left and right images can be moved individually. This allows the operator to view the vehicle's surroundings simultaneously from different directions. In addition to the bird's-eye view image, multi-view allows the operator to display two fixed-view images from among the front, rear, left, and right mounted cameras. This way, the operator can keep constant watch on fixed directions while getting a bird's-eye view of the vehicle's surroundings (**Figure 5**).

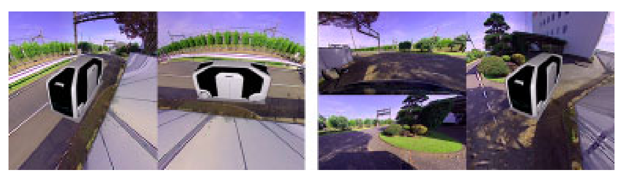

Figure 5. Example of Dual View (Left) and Multi-View (Right)

(4) AI Operation

FlyingView can be installed with an AI engine/ application developed according to the detection target and operating environment as a container application. This way, the AI can constantly monitor the pre-combined front, rear, left, and right camera images. A specific example of AI operation is described in "Operation Example of AI" presented later in the article.

Specifications

Table 1 shows the main specifications of FlyingView. The main unit is available in four models, which differ depending on the type of wireless communication (Wireless LAN or LTE*2)) and whether or not the unit is VPU-equipped.

*2) LTE is a registered trademark of the European Telecommunications Standards Institute (ETSI).

The units are equipped with four camera inputs for the front, rear, left, and right cameras required for generating the bird's-eye view images. For real-time distribution of the high-definition images, H.264 compression standard has been adopted. In terms of environmental resistance, IP protection class (IP66), Automotive Electromagnetic Compatibility (EMC) Standard compliance (equivalent to ECE-R10), and operating temperature range of -30°C to 60°C (-20°C to 60°C for VPU-equipped modes) have been achieved.

Table 1. Main Unit Specifications

		YS6010A	YS6010B	YS6010C	YS6010D	Remarks
Image	Camera Input	4ch (*Internally supports up to 6ch)				Dedicated cameras (HD)
	Monitor Output	1ch				Dedicated monitor (HDMI)
Compression	H.264	1ch				Dedicated console app
Communication	LAN	10/100/1000Base-T				
	WLAN	Yes	No	Yes	No	802.11b/a/g/n/ac/ 2x2
	LTE	No	Yes	No	Yes	
Recording	Bird's-Eye View Image Recording	Yes				Saved to microSD card
	Free Viewpoint Recording	Yes				Saved to internal SSD
VPU		No		Yes		Intel Movidius Myriad X
Others	Mic Input	Yes				USB mic
	GPS	No	Yes	No	Yes	LTE model only
	Contact Input	3ch				For camera switching
General Specifications	Dimensions	340mm(W) x 230mm(D) x 80mm(H)				Excludes protrusions such as antennas
	Operating Range	-30°C to +60°C		-20°C to +60°C		
	Power Supply	12 / 24 VDC				
	Power Consumption	45W		51W		Including cameras
	Weight	Approximately 6kg				
	Moisture/Dust Resistance	IP66 (with the unit's waterproof cover closed)				

Main Unit

The main unit consists of (1) a main processor that distributes the bird's-eye view images and performs AI processing, (2) a sub-processor that generates the bird's-eye view images, and (3) an AI processor that performs AI operations such as image recognition (**Figure 6**).

The main processor is an Intel Atom® *3) processor (E3950) and the AI processor is an Intel® Movidius™ Myriad™ X VPU. Similar to the AE2100, the AI engine/application developed according to the detection target and operating environment can be installed as a container application, therefore configuration is easily customizable according to the customer's needs.

A large-capacity FPGA (Zynq® UltraScale+™ MPSoC*4)) with XILINX ARM® *5) cores has been adopted as the sub-processor enabling the generation of the bird's-eye view images in real-time and with low power consumption. For the process of generating the bird's-eye view images from the four camera images, Xylon's multi-camera image stitching engine is installed with customization for FlyingView.

H.264 image compression and distribution of the bird's-eye view images are performed with the main processor.

A large-capacity/low-delay image transfer function using DMA was implemented between the main processor and the sub-processor. This makes it possible to distribute real-time images and record the images from the four cameras all at the same time, or perform the playback function while recording the four camera images.

Furthermore, the unit is configured for addition of two more optional cameras, which can be used when expanding functions in the future.

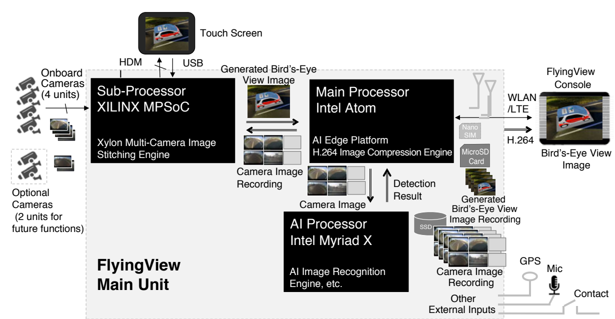

Figure 6. FlyingView Main Unit Configuration

Operation Example of AI

As an operation example of the AI, detection of people in images taken with the FlyingView cameras and external transmission of the results are introduced.

(1) Software Architecture

Figure 7 shows the software architecture that enables the AI operation. The basic configuration is the same as the AE2100, and the AI engine and AI application can be installed on a standard container running on top of Docker® *6). In this example, the following AI engine and AI application were prepared.

*3) Intel, Intel Atom, Movidius, Myriad and OpenVINO are trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
*4) Zynq and UltraScale+ are trademarks or registered trademarks of XILINX in the United States and other countries.
*5) ARM is a registered trademark of Arm Limited.
*6) Docker is a registered trademark of Docker, Inc.

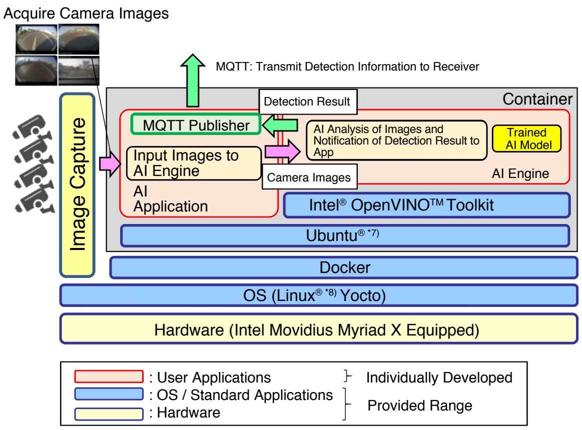

Figure 7. FlyingView Software Architecture

1) AI Engine

AI engine analyzes the pre-combined front, rear, left, and right camera images input from the AI application described below, and provides the coordinates of the detected people to the AI application. Detection is performed based on a skeleton estimation model called Human Pose Estimation (HPE). HPE detects human skeletal structures such as hands, elbows, shoulders, necks, and hip joints, and returns the coordinates of those detections. The AI model used in the example was trained using images from a FlyingView-equipped bus traveling within a factory premise.

2) AI Application

AI application acquires camera images from FlyingView and inputs the images into the AI engine. It also transmits the coordinates of the human skeletal structures that the AI engine detected to the console using the MQTT (Message Queuing Telemetry Transport) protocol. In the example, the coordinates of the detected human skeletal structures and the camera images were transmitted.

(2) AI Console

The results were confirmed by displaying the coordinates of the detected human skeletal structures and the camera images received from the AI application onto the console screen (**Figure 8**). The skeletal structures are superimposed on the people appearing in the camera images, and it shows that people can be correctly detected.

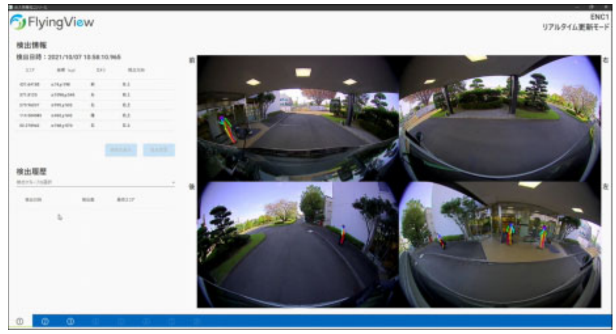

Figure 8. Screen Display of People Detection Results

Conclusion

This article explained the features, functions, and configurations of the remote monitoring system “FlyingView,” which with its “human eyes” and “AI eyes” prevents oversight of happenings that need to be seen.

FlyingView generates and distributes free bird’s-eye-viewpoint images in real-time enabling the system user to intuitively grasp situations surrounding a vehicle, etc. from a remote location. As an AI edge platform, FlyingView can be installed with AI engine/application developed according to the detection target or operating environment. This makes it possible to monitor the surrounding with “AI eyes.”

Future work will emphasize coordination between the “human eyes” and “AI eyes” to make advanced remote operation of vehicles, ships, robots, etc. possible, and thereby achieve OKI’s goal of “Delivering OK! to your life.” ◆◆

References

1) Takamitsu Shimada: Realizing High-Speed Deep Learning Inference with AI Edge Computer “AE2100,” OKI Technical Review, Issue 234, Vol.86 No.2, December 2019 https://www.oki.com/en/otr/2019/n234/pdf/otr-234-R05.pdf

Authors

Kazuki Kobayashi, IoT Systems Department, IoT Platform Division, Solution Systems Business Group

Masashi Yuzue, IoT Systems Department, IoT Platform Division, Solution Systems Business Group

*7) Ubuntu is a registered trademark of Canonical Limited.
*8) Linux is a registered trademark of Linus Torvalds in the United States and other countries.

TIPS【Glossary】

IP (International Protection)
Standardized waterproof and dustproof performance of electronic devices. The IP66 level of protection is defined as follows.
Dustproof protection: Complete protection from dust.
Water protection: Protection from powerful jets of water.

ECE-R10
International standard issued by the United Nations for in-vehicle immunity.

VPU (Vision Processing Unit)
A semiconductor chip that performs image processing. It is used to perform high-speed calculations required at the AI edge.

Zynq UltraScale+ MPSoC
SoC (System-on-a-Chip) provided by XILINX. Two processors, an application processing unit and a real-time processing unit, are installed on one SoC. Multiple cores installed in each processor enhance functionality.

FPGA (Field Programmable Gate Array)
An integrated circuit that can be freely programmed by the user.

DMA (Direct Memory Access)
One of the data transfer methods in a computer system. High-speed data transfer can be performed by bypassing the CPU and directly transferring data between peripheral devices and main memory (RAM).

MQTT (Message Queuing Telemetry Transport)
A simple Publish/Subscribe asynchronous, bidirectional protocol that runs over TCP/IP. One of the ideal network protocols for IoT due to its lightness and flexibility.

Container
A logical partition (container) created on the host OS into which libraries required to operate the application are packaged together as one. It is lightweight and operates at high speed since it requires less overhead than server virtualization.

Docker
Typical open source software for providing containers.