Edge Media Server for Real-Time 4K Video Streaming with Multiple 5G-Enabled Drones

Jae-Kwan Yun
DNA+Drone Platform Research Center
Electronics and Telecommunications Research Institute
Daejeon, South Korea
jkyun@etri.re.kr

MoonSoo Lee
DNA+Drone Platform Research Center
Electronics and Telecommunications Research Institute
Daejeon, South Korea
mslee@etri.re.kr

ChaeDeok Lim
Air Mobility Research Division
Electronics and Telecommunications Research Institute
Daejeon, South Korea
cdlim@etri.re.kr

Abstract— ***Recently, there has been a significant surge in the development of services utilizing multiple drones, particularly in domains such as detect and rescue operations. These advanced services leverage the capabilities of the high-speed 5G network to capture 4K videos from drones. Once these videos are sent through the MEC (Mobile Edge Computing) server, they are utilized for functions such as object detection, tailored to particular needs. To achieve this, the development of a media server capable of receiving, storing, and relaying 20 or more channels of 4K videos along with geometry information, is needed. During flight, over 20 drones simultaneously transmit data, including 4K videos at 8-20MBps bandwidth and 30 FPS(Frames Per Second), along with metadata containing sensor and gimbal information. This data is meticulously synchronized and formatted into MPEG-2 TS before being sent to a media server. Once transmitted, the media server receives and relays the videos to connected data servers, BM(Business Model) monitoring servers, and AI servers. These servers perform tasks such as storing, monitoring, and AI analyzing, enhancing the efficiency and relevance of the transmitted data.***

In a typical commercial 5G network, an upload speed of 100Mbps is common. However, in the case of the private "e-um 5G" network, upload speeds of up to 300Mbps or more are achievable. This enhanced upload capability enables up to 20 drones to transmit high-resolution video simultaneously. The media server developed in this paper is designed to receive and relay drone videos in the "e-um 5G" environment. To validate its capabilities, the media server was tested in a simulated environment with up to 50 drones connected and operational.

In this paper, the edge media server processes 4K videos and metadata transmitted from drones and relays them for display on the web through various protocols. Furthermore, the GPS information extracted from the streamed video on the web is utilized to annotate the locations covered by drones on digital mapping platforms like Cesium and VWorld. The system demonstrates the capability to present four video streams along with their associated metadata within a single web page.

Keywords—Multi-drone Media Server, 5G network, Drones, e-um 5G, Low Latency

I. INTRODUCTION

In recent times, there is a growing demand for various services utilizing 5G technology[1]. Specifically, the development of real-time video acquisition using drones and the application of AI technology to enhance accuracy and processing speed are in progress[2,3]. The advantages of 5G technology include its ability to transmit and receive 4K videos wirelessly compared to the traditional LTE network[4,5]. It also supports MEC (Mobile Edge Computing), allowing captured videos to be processed within the MEC without direct transmission to a central server. This reduces the cost of high-resolution video transmission and facilitates data exchange between MEC-equipped mobile base stations.

This paper presents the design and implementation of a media server for streaming 4K videos from N (N≤50) drones, intended for use in a missing person search service involving drones deployed on mobile 5G base stations. The media server's relay focuses on a range of inputs, including video footage obtained from cameras installed on drones, metadata derived from sensors and gimbals, and static images combined with corresponding metadata. Moreover, considerations for mobile browser display in dynamic environments are imperative. After the drone flight, stored videos are utilized both for playback using VoD(Video On Demand) and as input for AI analysis.

In contrast to traditional CCTV surveillance networks, which rely on continuous power and use closed networks for video transmission, drone networks utilizing commercial 5G or private "e-um 5G" networks involve diverse terminal connections within short flight times. Therefore, a CDN-based live broadcast approach is more suitable. To facilitate this, our paper supports protocols like 5G-based RTSP(Real-Time Streaming Protocol), RTMP(Real-Time Messaging Protocol), HLS(HTTP Live Streaming), FLV(Flash Video), Dash(Dynamic Adaptive Streaming over HTTP), WebRTC(Web Real-Time Communication), and MPEG-2 TS. The structure of this paper includes a related works in chapter 2, followed by the service scenario for the edge media server in chapter 3, system design in chapter 4, system implementation details in chapter 5, and concluding remarks in chapter 6.

II. RELATED WORKS

In this section, the integration of the media server in this paper with the DNA+Drone platform is explained, along with an elucidation of the KLV(Key Length Value) standard employed for acquiring metadata.

A. DNA+Drone Platform

The DNA+Drone platform integrates D (Data), N (Network), and A (Artificial Intelligence) with drones to achieve real-time data processing, enabling BVLOS(Beyond-Visual-Line-Of-Sight) operations, autonomous flight, remote control, and service creation. This initiative aims to establish a robust drone services ecosystem, focusing on the development of the following technologies:

This research was supported by DNA+Drone Technology Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. NRF-2020M3C1C2A01080819)”

- Operating DNA+Drone platform for AI-based drone data learning and validation using 5G.
- Developing a framework for collecting large-scale drone sensor data specific to validation services, along with standard data modeling techniques.
- Developing a 5G-enabled drone system and remote operational system that allows real-time remote autonomous operation of multiple drones using the 5G mobile communication network.
- Developing and openly sharing core AI technologies and software for services involving large-scale data from drone clusters, including situation comprehension, autonomous control, cluster collaboration, and emergency response.

B. *KLV Metadata*

KLV metadata is a structured format in UAV and imaging devices' video streams, encapsulating essential video-related information. The abbreviation signifies the trio of Key, Length, and Value, which constitute the fundamental elements of this structure. In video data, KLV metadata is commonly used to provide frame-specific details like subjects, coordinates, timestamps, and more. The Key functions as a unique identifier within KLV metadata, such as "Image Longitude" or "Timestamp." Length signifies the Value section extent, indicating data volume. This element is crucial for pinpointing metadata termination. The Value section holds actual information, like GPS coordinates or temporal data. KLV metadata is used in imaging, video analysis, GIS, surveillance, and security, serving as a vital tool for efficient video data management and interpretation. The paper refers to the MISB ST 0601 [6] standard for metadata of unmanned mobile platforms and proposes a minimal set for sensor models, defined in MISB ST 0902 [7].

III. The Service Scenario for the Edge Media Server

In this section, we describe a scenario involving the detection of people or objects using video streams transmitted by multiple drones during flight, within a commercial 5G or private "e-um 5G" network environment. Figure 1 show the service scenarios for the edge media server.

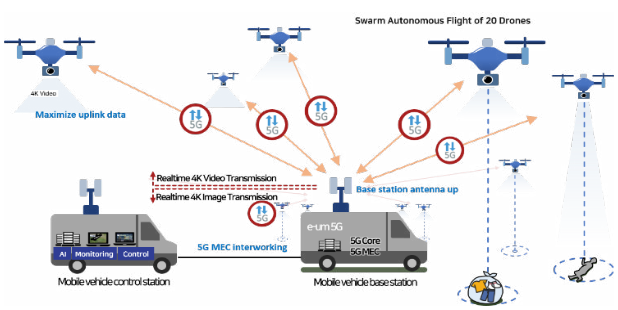

Fig. 1. The core service scenario for locating people or object using multiple drones within a commercial or private "e-um" 5G

First, real-time video from the 4K camera attached to the drone, along with a capture board, is encoded into H.264/H.265 and transmitted via RTP. The video receiver module of the media server receives the video and converts it into LIVE or VoD format according to the client's requirements. The functions of the media server using multiple drones are described below.

- (4K Video Transmission Capability): Using the wide bandwidth of 5G, the transmitted video from drones must be in 4K (3840x2160) at 30 FPS (Frames Per Second). Moreover, to enhance object recognition accuracy, a bitrate of 8-20Mbps needs to be provided. The media server should support 20 to 50 channels for transmission.
- (4K Video Transmission with Included Metadata) : Metadata collected from sensors and gimbals attached to drones is encapsulated in MPEG-2 TS format in KLV binaries. The MPEG2-TS formatted stream is received by the media server, which then displays it on web-connected clients. The KLV metadata should be synchronized with the video and displayed on web-based maps (Cesium, VWorld).
- (Multiple Protocols) : CDN utilize various low-latency protocols to ensure efficient and fast content delivery to users. WebRTC is particularly suitable for interactive applications like video conferencing and live streaming, QUIC (Quick UDP Internet Connections) is designed to provide faster and more secure data transfer compared to traditional TCP for web applications. SRT(Secure Reliable Transport) is an open-source protocol developed for secure and reliable video transport over the internet. HLS is commonly used in CDNs for adaptive streaming. MPEG-DASH supports low-latency streaming by using shorter segment durations and providing more efficient delivery.
- (Remote Control) : In the context of a drone connected to a 5G network, the MC (Mission Controller) cannot serve as a server. As a result, it is imperative to establish remote connectivity and control procedures once the drone commences its flight. Additionally, the media server should have the capability to disseminate the drone's network status information upon the 5G network. This enables real-time adaptations, such as modifying bandwidth or resolution, and potentially changing to alternative media servers contingent on the quality of the network.

IV. Design of the Edge Media Server

In this chapter, we provide an in-depth explanation of the design considerations behind the edge media server.

A. *Drone connection and remote control function*

Due to stringent security policies imposed by commercial 5G networks, drones connected via tethering to 5G devices cannot become servers. For this reason, drones transmit flight control information and 4K video streams to a separate server with a fixed IP. Initially, upon drone activation and 5G activation, the drone connects to a CMS(Connection Management Server) to register its unique IP and drone ID. Subsequently, it connects to a FC (Flight Control) relay server capable of relaying FC data and periodically sends mavlink data. As show in figure 2, the drone's MC then launches a streaming program for 4K streaming, simultaneously initiating archiving of the 4K video to local storage. While archiving is also stored on the drone during streaming, the media server receiving the stream also performs storage. Additionally, a remote control program is connected to address potential issues that may arise during drone flight, allowing real-time monitoring of the overall stream status and individual video FPS. Furthermore, if the 5G connection experiences fluctuation causing disruptions in transmission during streaming, the remote control program can dynamically adjust bandwidth or switch to alternate streaming relay servers.

B. 4K Video Transmission

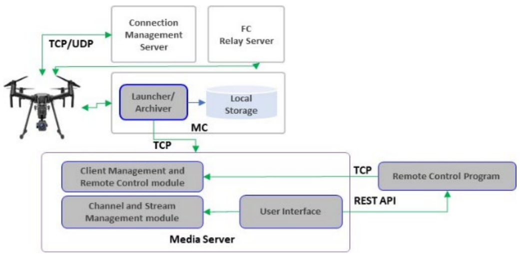

Fig. 2 The data flow of drone connection and remote control function

Individual drones operating in a 5G environment can transmit either video alone or video with included metadata. For 4K video streams, you have the option to use UDP for low-latency transmission or TCP to minimize data loss. When utilizing UDP transmission, protocols such as RTP or SRT are used, whereas TCP transmission involves utilizing RTMP to push the video content. As show in figure 3, the edge media server can establish a one-to-one connection with a drone or function as an intermediary in rendezvous mode, relaying the transmission module of the drone to different modules.

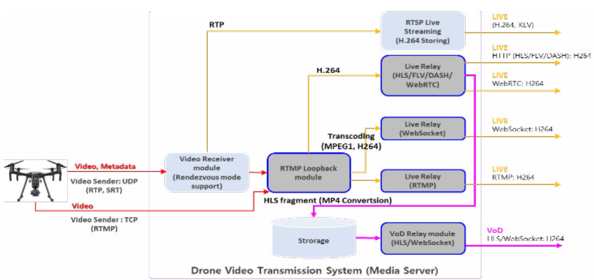

Fig. 3. The data flow of 4K video transmission

In Rendezvous mode, both the transmission and receiving modules connect to the media server before streaming takes place. This mode proves useful when an external operator, like a 5G mobile control station in a car, receives drone stream data from an on-site media server. For displaying 4K streams to users, you can leverage either web or app solutions. In the case of app-based display, you would convert UDP-transmitted H.264 video into RTMP or RTSP formats for output. Considering that many formats are supported on the web, including those compatible with RTMP input, this system can relay UDP-transmitted H.264 video via RTMP loopback, subsequently relaying it using HLS, FLV, DASH, and WebRTC. During stream transmission, local storage is used, and MP4-converted video files are played back through a VoD approach.

C. 4K Video with Metadata Transmission

The proposed edge media server incorporates an AI server, facilitating the seamless transmission of real-time streaming 4K videos as input to the AI server. Within the AI server, functions such as missing person detection, crop change detection, and water boundary detection are performed using the transmitted videos. To achieve this, it is crucial to synchronize data from the drone's gimbals and sensors with the videos, relay them, and present metadata alongside 4K videos on screens through web and app interfaces. Figure 4 show the data flow or 4K video with metadata transmission.

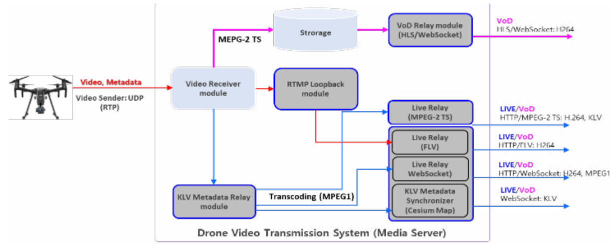

Fig. 4. The data flow of 4K video with KLV metadata transmission

The inserted metadata is converted into a byte stream using KLV encoding and input into an MPEG-2 TS container. The media server receives the transmitted MPEG-2 TS stream from the drone and archives it in Local Storage for VoD streaming. When processing videos in an app, video frames and KLV metadata are sequentially separated from the received MPEG-2 TS and displayed on the screen. For web-based video processing, two methods can be used: one employing JavaScript on a webpage to display videos and KLV metadata in sync, and the other involving middleware to convert videos into MPEG-1 Video or FLV format before displaying them on a webpage in sync with KLV metadata. While directly processing MPEG-2 TS on the web is a more advanced method, it relies on browser support and is employed selectively due to this dependency.

V. IMPLENTATION OF THE EDGE MEDIA SERVER

In this section, the results are presented for video and KLV metadata, displaying the result of the 4K media transmission, and 4K video and KLV metadata map mapping.

A. Alignment of Video and KLV Metadata

The KLV metadata utilized in the implementation of this paper comprises 25 ~ 26 tags, organized as binary data of 240 bytes. In the case of a video within an MPEG-2 TS stream running at 30 FPS, 30 frames and 30 sets of KLV metadata are consecutively extracted every second. Figure 1 illustrates the display of received KLV metadata alongside the video content on a web interface. As show in figure 5, tags 4 to 7 represent current drone flight information, heading, pitch, and roll angles. Tags 13 and 14 indicate the GPS coordinates of the drone's current flight, while tags 18 to 20 represent the pitch, yaw, and roll values of the gimbal. Tags 82 to 89 denote the coordinates of the camera's viewport capturing the current drone's imagery. These coordinates are employed to mark the red rectangle on the right side of the illustration.

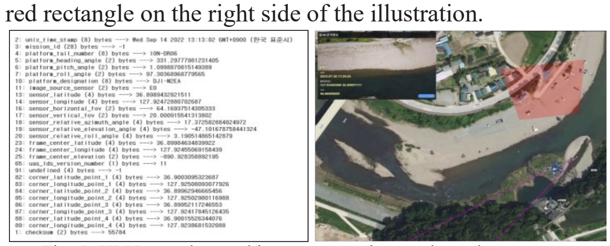

Fig. 5. KLV metadata and its representation on the web page

B. Displaying the Outcome of the 4K Media Transmission

Video received via RTP is relayed through RTSP and RTMP, adapting to the chosen web display method. For instance, video input via RTP is pushed to an RTMP local server. Users can access the video for display either through an application-like interface connecting to the RTMP server or through web formats like FLV, HLS, Dash, or WebRTC. Each approach may introduce variations in video quality and latency. If the browser supports MSE(Media Source Extension), streaming benefits from the browser's assistance in handling and displaying 4K streams. However, in cases where MSE is not supported, a straightforward MPEG-1 format is employed to display the stream on the web. FLV offers good video quality but can lead to transmission delays, whereas WebRTC features lower latency but may experience interruptions with high bit rates. Figure 6 show simulated drone flight and video streaming.

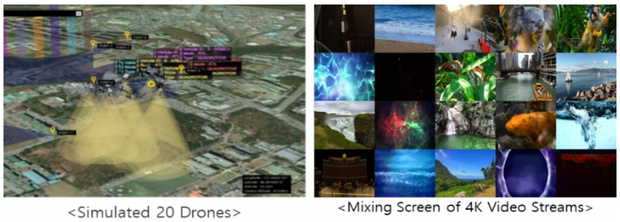

Fig. 6. Displaying the Outcome Screen of the 20 Drones Simulation and Merged View of 20 Channels of 4K Video Streams

C. 4K video and KLV metadata Map Mapping

For videos with embedded KLV data, the media server uses FLV format for web-based video display. KLV metadata is separated from the video on the media server and sent to the web page via WebSocket, with JavaScript employed for metadata demuxing. KLV metadata is generated at the same frequency as video frames, and a CustomEvent approach is used to obtain and showcase the data on a map within the browser's callback function. The browser's performance dictates the number of CustomEvents generated per second, with Chrome supporting up to five occurrences per second. Figure 7 show the mapping resuls of 4K video and KLV metadata on the web map.

Fig. 7. Mapping Results of 4K Video and KLV Metadata on the Web Page

Furthermore, depending on the business model involving drones, there might be a requirement to display 4K videos and metadata received from multiple drones on a single webpage. Figure 8 shows the display of 4K videos and metadata simultaneously transmitted from four drones.

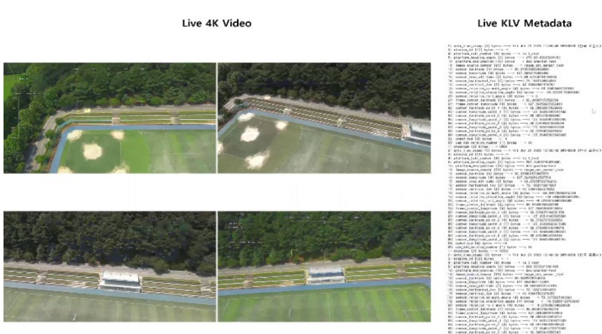

Fig. 8. Display Results of 4K Videos and KLV Metadata from Four Drones

The video metadata transmitted from the drone is not only utilized for real-time monitoring in a live environment, but also serves as input for real-time AI analysis. In Figure 9, the outcome of utilizing 4K video metadata as input for object tracking in a web environment is displayed, along with a screen showing the execution of replay using stored MPEG-2 TS files in Local Storage.

Fig. 9. User Interface for AI with 4K Video Input and VoD Display

In Figure 1, it demonstrates the process where the media server receives 4K video metadata transmitted from the drone and displays it on a web page. Simultaneously, an AI analysis server detects objects and displays the results.

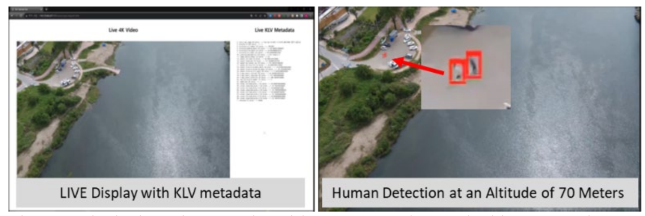

Fig. 10. Displaying Live Results with KLV Metadata and Object Detection

VI. CONCLUSION

The rising demand for services utilizing multiple drones, particularly in detection and rescue operations, leverages the 5G network's speed to transmit multiple 4K drone videos. A capable media server is pivotal, receiving, storing, and relaying numerous 4K video channels and geometry data. Over 20 synchronized drones transmit data, including 4K videos, metadata, and sensor info, managed by the media server. Data efficiency and relevance are boosted through relay to data servers, BM monitoring, and AI servers.

The paper focuses on the media server processing 4K drone videos and metadata, relaying them using protocols like WebRTC, Dash, HLS, WebSocket, and FLV. Extracted GPS info marks drone locations on platforms like Cesium, displaying four video streams with metadata on one web page.

REFERENCES

[1] D. Ryoo, K. Kim, and C. Lim, "Implementation of Real-Time Facility Inspection Service System Based on a Drone," International Conference on Information and Communication Technology Convergence (ICTC). pp. 2377-2379, 2022.

[2] J. Jin, et al., “Design of UAV video and control signal real-time transmission system based on 5G network,” IEEE 16th Conference on Industrial Electronics and Applications (ICIEA), pp. 533-537, 2021.

[3] S. Kim, et al. “A Study of real-Time 4K drone images visualization to rescue for missing people base on web,” International Conference on Information and Communication Technology Convergence (ICTC), pp. 1594-1596, 2022.

[4] F. Ahmed, J. C. Mohanta, A. Keshari and P. S. Yadav, “Recent advances in unmanned aerial vehicles: a review,” Arabian Journal for Science and Engineering, pp. 7963-7984, 2022.

[5] H. Koumaras, et al. “5G-enabled UAVs with command and control software component at the edge for supporting energy efficient opportunistic networks,” Energies, 14.5: 1480, 2021.

[6] MISB ST 0601.16a UAS Datalink Local Set, 1 May, 2020.

[7] MISB ST 0902.8 Motion Imagery Sensor Minimum Metadata Set, 1 Nov, 2018.