Remote operation of two construction machines from a location 900 km away using a single Starlink Mini and 10 HD video streaming

Yasuyuki Nakajima[1], Yasukazu Hontama[1], Daisuke Endo[2], Genki Yamauchi[2], Takeshi Hashimoto[2], Yasuyuki Jitsuta[3], Hitoshi Itoh[4], and Masayuki Okamura[4]

[1]Hytec Inter Co. Ltd., Japan
[2] Public Works Research Institute, Japan
[3] Jitsuta Chugoku Co. Ltd., Japan
[4]Chudenko Corporation, Japan

y-nakajima@hytec.co.jp, hontama@hytec.co.jp, endou-d177cl@pwri.go.jp, yamauchi-g573bs@pwri.go.jp, t-hashimoto@pwri.go.jp, y.jitsuta@jtch.co.jp, h.itoh@chudenko.co.jp, m.okamura@chudenko.co.jp

Abstract –

Expectations are growing for low-orbit satellite communication network to enable remote operation of construction machinery in mountainous areas and disaster recovery sites where public communication infrastructure is difficult to establish. However, transmitting multiple camera feeds for remote operation of multiple machines poses challenges, as bandwidth constraints and packet loss make stable video transmission difficult. Therefore, we developed BAERT® (Bandwidth Adaptive and Error Resilient video Transmission) technology to solve the above problems. The developed method adapts to available network bandwidth and enhances error resilience to enable video transmission. As a proof-of-concept experiment, we successfully transmitted video from 10 HD cameras using a single Starlink Mini satellite and remotely operated two construction machines.

Keywords –

Remote construction, Error resilient transmission, Video encoder, Bandwidth adaptation, Starlink

1 Introduction

Japan's working-age population is estimated to decrease by 20% by 2040 compared to 2020. The Ministry of Land, Infrastructure, Transport and Tourism's "i-Construction 2.0" plan, formulated in 2024, aims for a 30% productivity increase through labor-saving measures achieved by automating construction processes [1]. In this scenario, remote construction serves as a crucial step toward subsequent automation and autonomy. Furthermore, under the i-construction 2.0 initiative, it is anticipated that in the future, autonomous and automated construction methods will be used in combination with remote construction, with each approach employed where most appropriate. For this reason, efficient remote construction is expected to become a key technology in the years to come.

Remote construction is broadly categorized into three methods [2, 3]. The first is the "direct visual operation" method [2], where construction machinery is being remotely operated using a remote-control device from a location 30 to 50 meters away. The second is "remote operation using visual monitoring," where operation is performed from a remote-control room located up to approximately 300 meters away, using images from cameras mounted on the construction machinery or using images from fixed cameras in the field. The third is "ultra-long-distance remote operation using visual monitoring," [3] which applies when the operating distance from the remote-control room to the construction machinery exceeds 300 meters.

In recent years, the proliferation of broadband networks, advances in high-compression video encoders, and the low-cost high-resolution video cameras have drawn attention to "ultra-long-distance operation via visual monitoring," which utilizes the Internet to operate construction machinery from remote locations. This approach eliminates location constraints for operators, reducing travel costs and time to sites while addressing skilled labor shortages. Its adoption is also expected to help achieve the new opportunities for construction workers: good pay, time off, and a sense of hope [1]. Furthermore, regarding networks, there is growing expectation for the use of Low Earth Orbit (LEO) satellites like Starlink, which enables significantly higher-capacity communications compared to conventional geostationary orbit (GEO) satellites, in areas where existing fiber-optic networks or 5G networks are difficult to utilize, such as mountainous regions and disaster recovery sites.

However, construction machinery must be equipped with approximately four high-definition video cameras covering front, rear, left, and right for safety checks, and additional one or two bird's-eye view cameras are required to grasp the site conditions [4]. Regarding the network, previous experiments have achieved throughputs of approximately 30 Mbps for uplink (from terminal to base station) on carrier LTE and around 10 Mbps on Starlink [4]. However, with a video encoding rate of 2–3 Mbps per camera, transmitting video from 5–6 HD cameras, including the bird's-eye view cameras, presents a significant hurdle.

This paper demonstrates that, in video transmission, by controlling the encoded bit rate within the network bandwidth and enhancing resilience against transmission errors, it is possible to transmit video from multiple cameras even in environments where transmission errors occur within limited bandwidth. To achieve this objective, we developed BAERT (Bandwidth Adaptive and Error Resilient video Transmission), which integrates the bitrate control method with error resilience both described in [4], and implemented it in the 4-channel HD encoder LVRC-4000. Regarding bitrate control, we have demonstrated encoding at 3 Mbps; however, we also confirmed that even in the high-compression mode of 1 Mbps, bitrate fluctuations were low and could be kept below 1 Mbps. Furthermore, as described in Section 2.1, conventional MPEG video encoders use IBP or IP structures, which resulted in significant frame-to-frame variation in coding volume. BAERT resolved this issue by applying Intra Slice encoding within an IP structure.

In a proof-of-concept experiment, it was confirmed that remote operation of construction machinery was achievable by transmitting a total of 10 high-definition video streams—eight streams from video cameras mounted on two construction machines and two overhead video camera streams—from an experimental field in Tsukuba City, Ibaraki Prefecture, to Numata Town, Hokkaido, 900 km away, using a single Starlink Mini satellite. To the best of our knowledge, this is the first attempt of its kind in the world.

Subsequent chapters describe the developed video transmission technology, the configuration of the demonstration experiment, and the experimental results.

2 Remote Operation of Multiple Construction Machines in Diverse Wireless Environments

This chapter discusses video encoding techniques for limited network bandwidth and countermeasures against transmission errors as technologies necessary for operating multiple construction machines in various wireless environments.

For efficient remote operation of construction machinery, camera footage mounted on the machinery must be transmitted with low latency, high clarity, and stability. Regarding the network connecting the remote location to the operator's seat, there are two methods: directly connecting to the remote location via the Internet using a carrier LTE (or 5G) router mounted on the machinery; or using a local network such as Wi-Fi or private 5G at the construction site, then connecting to the remote location via fiber optic or satellite network. Regarding the former method, throughput of approximately 30 Mbps and transmission latency of around 50 ms were observed in the previous experiments [4], though these depend on coverage and congestion levels of the networks.

Regarding the latter, since the construction site network is closed network, it is possible to build a high-speed, low-latency environment with throughput exceeding 30–50 Mbps and latency ranging from several milliseconds to tens of milliseconds. On the other hand, if the connection beyond that point uses fiber-optic lines, a sufficiently high-speed, low-latency environment can be achieved. However, while LEO (Low Earth Orbit) satellite networks like Starlink can offer significantly higher capacity and lower latency compared to conventional GEO (Geostationary Earth Orbit) satellite networks, Starlink has shown uplink throughput of around 10 Mbps and latency of approximately 60 ms in the previous experiment [4].

Therefore, if low-bit-rate encoding can be used to transmit numerous high-definition video streams over such communication networks, it is expected that multiple construction machines can be remotely operated across diverse wireless communication environments.

The following section describes methods for video transmission over limited bandwidth and for error tolerance.

2.1 Video Streaming over Limited Network Bandwidth

Video transmission over limited bandwidth requires the ability to control the coding rate below a specified rate and stable coding quantities for each frame. The international video coding standard MPEG allows selection from several coding structures based on coding efficiency and coding delay requirements. Figure 1 shows the most common coding structure, the IBP structure [5]. In this case, the first frame is an I (Intra)-frame, encoded independently without using other frames. The next frame encoded is a P (Predictive)-frame, shown as the third frame in the figure. This frame performs motion compensation prediction based on information from the temporally preceding frame. The subsequent frame encoded is a B (Bi-directionally predictive)-frame, shown as the second frame in the figure. This frame performs motion compensation prediction using information from both temporally preceding and following frames. Compared to P-frames, B-frames can predict from more frames, resulting in higher coding efficiency. Note that the size of each frame's box represents the data volume in Figure 1.

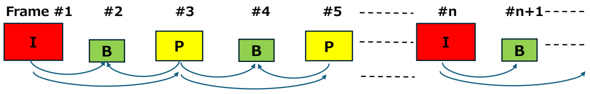


Figure1. MPEG Coding Structure（IBP Structure）

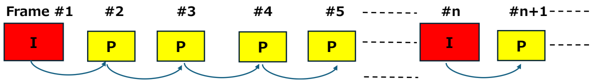


Figure 2. MPEG Coding Structure（IP Structure）

Next, Figure 2 shows the IP structure often used in low-latency mode. In this case, the B-frames present in Figure 1 do not exist, and the structure consists solely of I-frames and P-frames. Consequently, the coding efficiency of the IP structure is lower compared to the IBP structure. However, since B-frames cannot be encoded unless the preceding and following frames are encoded, both the encoder and decoder experience increased latency in the IBP structure compared to the IP structure.

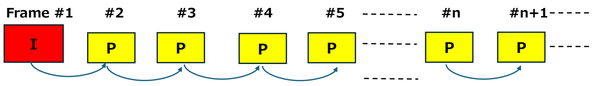


Figure 3. MPEG Coding Structure
（IP Structure + Intra Slice）

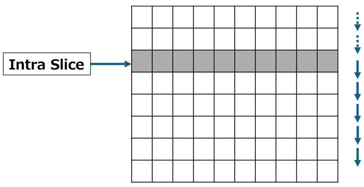


Figure 4. Intra Slice Mechanism

An example of an IP structure modification is shown in Figure 3. As can be seen from the figure, although it is an IP structure, there are no subsequent I-frames. In the cases of Figure 1 and Figure 2, I-frames are periodically inserted to refresh the video sequence. This allows recovery after the arrival of the next I-frame, even if a network disconnection occurs midway. In the case of Figure 3, the structure involves inserting intra slices, as shown in Figure 4, into each P-frame to constantly refresh a portion of the video sequence. In this case, a horizontal column of blocks within a given P-frame is encoded as a single slice, treated as intra-coded blocks that do not use any other frames. The next frame then encodes the block column below it as intra-coded. This enables encoding without periodically inserting I-frames, which leads to both lower latency and less data rate variation when compared with IBP structure.

Table 1. Comparison of MPEG coding performance

Coding structure	IBP	IP	IP＋Intra Slice
Coding efficiency	H	M	M
Data size variation	L	M	S
Coding delay	L	M	S

(H : High, M : Medium, L : Large, S : Small)

Table 1 compares the performance of these three coding structures. Due to the high interframe correlation in moving images, the IBP structure, which combines P-frames and B-frames for prediction, achieves the highest coding efficiency. Regarding the data size per frame, B-frames, which heavily utilize prediction from preceding and subsequent frames, have the smallest size, while I-frames have the largest. Consequently, the IBP structure exhibits the largest variation in the amount of data encoded per frame. Conversely, the intra slice structure, where only the first frame is an I-frame and all subsequent frames are P-frames, results in the smallest data variation. Regarding encoding delays, as mentioned earlier, the IBP structure incurs the largest delay. While the IP structure enables low-delay encoding, the intra slice structure, due to its minimal data variation, allows for consistently low and stable delay. These results indicate that the intra slice structure is optimal for achieving low encoding delay while minimizing data rate variation.

Furthermore, even with the same coding structure, to keep the data volume itself within a specified range, proper rate control is necessary to ensure the amount of coding required when quantizing the data in each coding block is appropriate.

2.2 Resilience for Transmission Error

Transmission errors can occur either on the network at the construction site or on the network between the construction site and a remote location. Regarding the former, packet loss may occur due to obstruction between the base station and construction machinery, caused by the movement of construction equipment or crane operations. Regarding the latter, packet loss may occur due to congestion on the transmission path to the remote location or noise in the wireless propagation. MPEG employs both spatial and temporal prediction techniques. Consequently, lost packets may cause partial image distortion, which may propagate spatially and temporally. In either case, error tolerance measures such as ARQ (Automatic Repeat reQuest), FEC (Forward Error Correction), and hybrid ARQ combining these methods [6] can be applied.

ARQ detects packet loss at the receiver and notifies the transmitter to request packet retransmission. If the retransmission arrives within the receiver's video decoding timing window, it can recover the lost packet and decode/playback normally. Therefore, a relatively large receive buffer is prepared to enable recovery even with a certain degree of packet retransmission delay. However, in cases where burst packet loss frequently occurs, decoding timing delays may occur, and retransmitted packets may fail to arrive in time for decoding, potentially causing frames to be dropped.

FEC involves transmitting data with error-correcting codes along with coded video data, and using this information for recovery when errors are detected at the receiver. Since it does not require feedback from the receiver to the transmitter, it is often used for one-way transmission such as broadcasting. However, increasing the amount of error-correcting code is necessary to achieve higher error correction capability. Regarding this overhead, when the amount of error correction code information is sufficient relative to the packet loss rate p (e.g., when the ratio $\varphi = k/N = 1.1p$, where k is the number of FEC packets and N is the total number of packets), packet transmission efficiency increases. However, when the amount of error correction code information is reduced (e.g., $\varphi = k/N = 0.9p$), packet loss cannot be fully corrected, leading to a significant decrease in transmission efficiency [7].

Hybrid ARQ consists of hierarchically transmitting FEC data during packet loss, resulting in a complex structure. For example, if packet loss is detected on the receiving end, it notifies the sending end. The sending end does not retransmit the packet but sends an error correction code instead. If an error is still detected and reported back to the sending end, the sending end sends an additional error correction code. The receiver then reconstructs the packet by combining all the received data. Therefore, it enables efficient error correction in terms of throughput [6] and is utilized as a wireless transmission protocol for 5G. Nevertheless, compared to Hybrid ARQ, which repeatedly retransmits data for burst errors, ARQ may offer higher recovery efficiency especially when feedback delay is large.

Therefore, ARQ is well-suited for enhancing error resilience with low latency and limited network bandwidth, as it can be implemented using relatively simple schemes. Furthermore, as mentioned earlier, MPEG undergoes periodic refreshes via intra-frames and intra-slices, enabling recovery during intra-coding even when packet loss occurs in bursts.

3 Remote Control Experiment System for Multiple Construction Machines

3.1 Experimental Setup

The transmission rate control and error resilience technologies described in Sections 2.1 and 2.2 were developed as BAERT (Bandwidth Adaptive and Error Resilient video Transmission) technology and implemented in the LVRC-4000 (Figure 5), capable of encoding four-channel video [4]. For the network connection from the remote location, Starlink was used.

While LEO network satellite has bandwidth limitations compared to optical networks or 5G, it is usable even in areas where broadband network deployment is difficult, such as mountainous regions. Regarding Starlink, in addition to the standard antenna (approximately A2 size), the compact Starlink Mini (approximately A4 size) is now available. Furthermore, since the introduction cost can be significantly reduced, Starlink Mini (Figure 6) was used in this experiment.


Figure 5. LVRC-4000


Figure 6. Starlink Mini

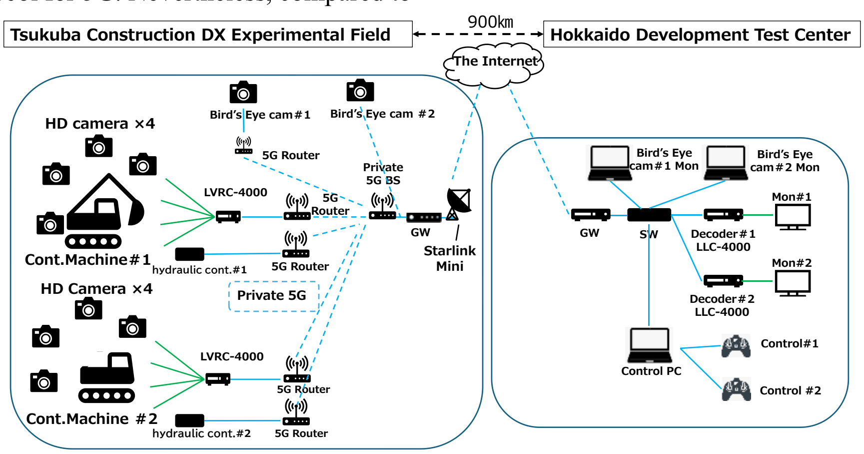


Figure 7. Experimental configuration

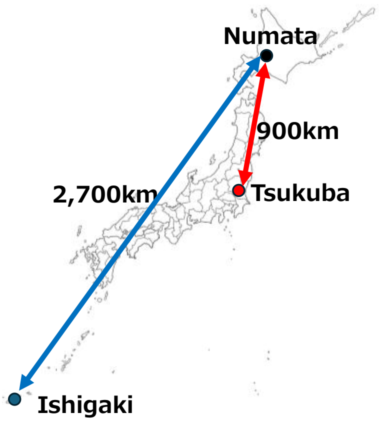


Figure 8. Location map of experiments

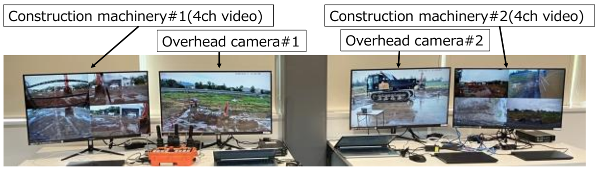


Figure 9. Ten channel video monitors at remote site

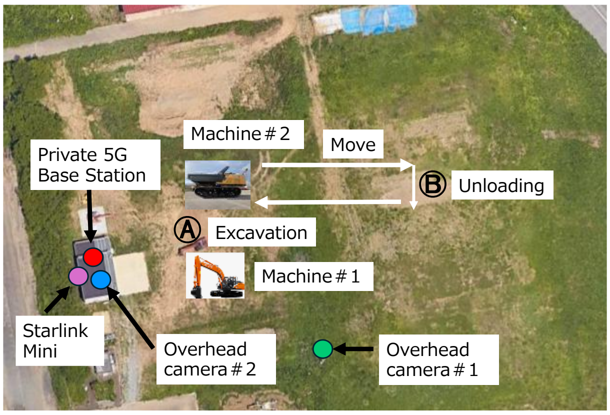


Figure 10. Experimental field

The construction site (Tsukuba Construction DX Experimental Field [8]) has two construction machines installed: Crawler Dump Truck: Morooka MST-110CR). Four full HD cameras (Omron STC-HD213DV) were mounted in front, rear, left, and right of each machine. Video was compressed using H.265/HEVC and transmitted via the LVRC-4000. Additionally, two optical 32x zoom PTZ full HD cameras (Hytech Inter HIC-SP200TX32A) were installed for bird's-eye views. Video encoders mounted on each bird's-eye view camera compressed and transmitted the video using H.265/HEVC. The overall configuration is shown in Figure 7. The combined eight channels of camera footage from the construction machinery and Bird's-Eye View #1 were transmitted via private 5G (Hytech Inter Compact private 5G Platform [9]), while Bird's-Eye View #2 was connected directly to Starlink Mini.

Remote operation was conducted from the Hytec Inter office (Hokkaido Development Test Center) in Numata Town, Hokkaido, approximately 900 km from the construction site as shown in Figure 8. An optical fiber connection was used for the Internet at the remote operation site. Four-channel video from each construction machine was displayed on individual monitors, while footage from two overhead cameras was shown on two separate monitors (Figure 9). Regarding the remote controllers used to operate the construction machinery, the backhoe was controlled using Hitachi Construction Machinery's S-RS-A, while the crawler dump truck was operated using a commercially available game controller. For both machines, the control signals were transmitted over IP using a protocol based on the common control signal currently under development at the Public Works Research Institute. The construction machinery control signals were transmitted over Starlink Mini and the private 5G network, along with the coded video data.

3.2 Experiment

First, we investigated performance metrics such as latency in a network-only configuration (private 5G + Starlink Mini + the Internet (fiber optic network)), excluding cameras, video codecs, and monitors, using the experimental setup shown in Figure 7. Next, we transmitted video from a total of eight cameras mounted on construction machinery and two overhead cameras, remotely operated two construction machines, and compared video quality between cases where BAERT technology was applied (to the construction machinery camera video) and cases where it was not applied. The construction machinery operations involved excavating soil with a backhoe, loading the soil onto a crawler dump truck (Figure 10 Ⓐ), and then having the crawler dump truck move to unload the soil (Figure 10 Ⓑ).

4 Experimental Results

4.1 End-to-End Network Performance

Table 2 and Figure 11 show the network performance when connecting the construction site (Tsukuba) using private 5G and Starlink Mini, and the remote operation site (Numata) using optical fiber. As evident from the table, the average throughput from the construction machinery achieved around 30 Mbps. Since private 5G typically offers throughput levels in the hundreds of Mbps, the overall network throughput depends on Starlink Mini's performance. Furthermore, while Starlink Mini specifications list an uplink speed of approximately 10 – 30 Mbps maximum [10], the experiment yielded values within this figure.

Table 2. End-to-End Network Performance in between Tsukuba and Numata (900km)

Conditions	Throughput (Mbps)				Delay(ms)↓				Jitter (ms)
	AVG	Max	Min	σ	AVG	Max	Min	σ	
Tsukuba->Numata	28.3	29.8	21.4	2.8	41.7	179	28.5	13.3	1.5
Numata->Tsukuba	67.6	71.3	42.4	6.3	41.7	179	28.5	13.3	1.5

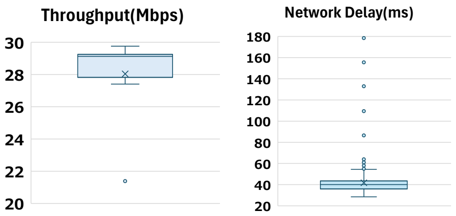


Figure 11. Network throughput and delay (From Tsukuba to Numata)

Although the average latency is around 40 ms, the maximum-minimum variation of latency is six times as large, which can be observed in Figure 11. Since private 5G exhibits virtually no latency variation, it can be said that the variation is largely due to Starlink Mini. Here, jitter is calculated as the average of the absolute values of the delay differences between consecutive packets. Therefore, the jitter value of 1.5 ms listed in Table 2 indicates that the delay differences between consecutive packets are not very large, meaning it is unlikely that packet arrival will suddenly be significantly delayed, which could significantly impact video playback. However, as shown in Table 2, the maximum delay is 179 ms and the minimum is 29 ms. This suggests that packet arrival may gradually become delayed, potentially causing video playback to pause for a short time. We have observed this phenomenon in cases where the system lacks error resilience.

Figure 11 also shows that while the delay is generally stable around 40 ms, delays several times larger frequently occur, potentially significantly impacting video transmission quality. Note that latency and jitter are calculated from round-trip time (RTT) measurement data, so the table shows values converted to one-way levels (= RTT/2).

We conducted another end-to-end network performance test between Ishigaki Island, Okinawa and Numata, Hokkaido (approximately 2,700 km) as shown in Figure 8. For this test, Starlink Mini terminals were installed at each location. The test results are shown in Table 3 and Figure 12. Due to limitations of the network tester, standard deviation of latency is missing.

As evident from the table, using LEO (Low Earth Orbit) satellites at both ends resulted in an average latency increase of 10 ms compared to the results in Table 2. Compared to the transmission experiment from Tsukuba to Numata, the ground distance increases by approximately 1,800 km. The additional distance due to the Starlink connection on the Numata side is about 550 km (either uplink or downlink distance), resulting in a total increase of 2,350 km. Therefore, an additional delay of 10 ms can be considered reasonable.

Table 3. End-to-End Network Performance in between Ishigaki and Numata (2,700km)

Conditions	Throughput (Mbps)				Delay(ms)↓				Jitter (ms)
	AVG	Max	Min	σ	AVG	Max	Min	σ	
Ishigaki->Numata	46.1	78.1	7.2	21.3	51.0	83.5	32.0	NA	0.25
Numata->Ishigaki	44.0	72.6	9.2	15.2	51.0	83.5	32.0	NA	0.23

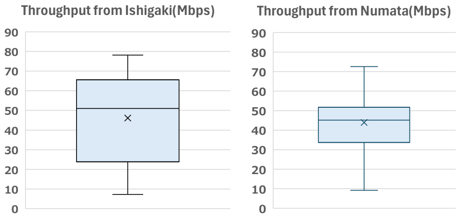
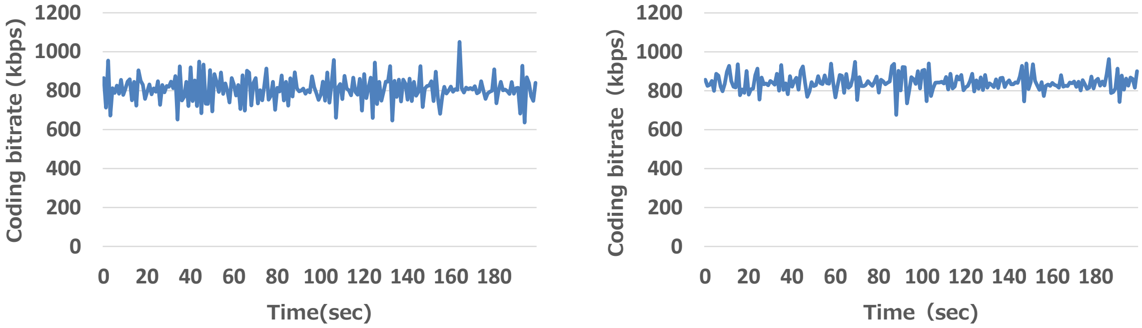


Figure 12. Network throughput when Starlink is installed at each location (Left: from Ishigaki, Right: from Numata)

Furthermore, since Starlink has significantly lower upload throughput than download throughput [10], comparing the upper row (upload to Starlink) and lower row (download from Starlink) in Table 2 reveals a significant throughput difference. In contrast, Table 3 and Figure 12 show almost the same throughput (about 45 Mbps) at each location because Starlink was used for uploads at both ends. Note that this figure exceeds Starlink's typical specifications, but throughput is believed to depend on the number of users simultaneously connected to the satellite. Therefore, it is assumed that the number of users was low at the time of this experiment.

4.2 Video Transmission Experiment

In the preliminary experiments using the configuration shown in Figure 7, when the average encoding rate for each camera video exceeded 1 Mbps, significant image distortion occurred due to the impact of delay variation. Therefore, the encoding rate for all camera images was set to 1 Mbps. The frame rate was set to 30 fps for the eight cameras mounted on construction machinery and 15 fps for the two overhead cameras. Figure 13(a) shows an example of encoding rate fluctuation with BAERT off (Crawler Dump Camera #2), and Figure 14(a) shows the image sample of actual video. Similarly, BAERT on is shown in Figure 13(b) and Figure 14(b). Note that the encoding structure with BAERT off uses an IP structure. This method is commonly used in low-latency encoding; we use BAERT off as the conventional method (baseline) and compare it with the proposed method (BAERT on).



(a) BAERT OFF (baseline) (b) BAERT ON

Figure13. Coding bitrate fluctuation

(a) BAERT OFF (baseline) (b) BAERT ON

Figure 14. Examples of received 4ch video

Table 4. Average and SD of coding bit rate

	Average bitrate	Standard deviation
BAERT off	807kbps	65.1kbps
BAERT on	842kbps	43.8kbps

Table 5. Dropped/Degraded segments/frames

BAERT On/Off	# of Deg/Drp Segments	Average segment size	Sigma	Total Deg/Drp frames	Deg/Drp ratio
OFF	33	51 frames	44 frames	1524	42%
ON	4	7.3 frames	1.9 frames	29	0.8%

The average coding rate and standard deviation are shown in Table 4. From Figure 13 and Table 4, it can be seen that enabling BAERT not only increases the average coding rate by about 5%, but also reduces the variation by more than 30%, enabling stable coding within the specified rate while the available bandwidth is being utilized as much as possible. Note that while this encoding rate control ensures that the coded data does not exceed the preset encoding rate, it does not automatically adapt to changes in network bandwidth. Consequently, if network bandwidth drops suddenly, the coded data size will exceed the available bandwidth; as a result, the decoder will only be able to play back the encoded data that arrives in time for playback, leading to dropped frames. Since such situations are anticipated in public 5G networks, we are currently investigating schemes that predict network bandwidth and adapt to it. However, in the Starlink network used in this study, while fluctuations in bandwidth were often observed, it never remained at a drastically reduced level, and we have confirmed that the proposed encoding control can be fully utilized. Additionally, while this depends on user preferences, a countermeasure such as using only the front camera in the event of a sudden drop in bandwidth could be considered.

Furthermore, comparing Figure 14(a) and (b), it is evident that with BAERT off, packet loss causes image degradation (red-circled areas) in multiple channels. Conversely, with BAERT enabled, timely packet retransmission ensures transmission without image quality degradation.

We investigated frame-by-frame video quality degradation and dropped frames caused by packet loss for 30-second of the received video shown in Figure 14, comparing results with BAERT enabled and disabled. The results are shown in Table 5. With BEART off, image quality degradation was observed 33 times across all four video channels, affecting 42% of the total frames. In contrast, while no image quality degradation was observed with BEART on, the number of dropped frames was 29, affecting only 0.8% of the total. This confirms a significant improvement from an objective standpoint.

(a) Remote operation site

(b) Construction field

Figure 15. Remote construction experiment

Regarding total delay from the construction site camera to the remote-control console monitor, it was 164.1 ms with BAERT off and 169.2 ms with BAERT on. Previous research indicates that end-to-end delays (from camera input to monitor output) below 300 ms are necessary for efficient remote construction operations [11]. This experiment achieved a latency well below this threshold, confirming that a single Starlink Mini satellite can transmit video from 10 high-definition cameras located 900 km away, enabling the operation of two construction machines. The remote operators reported perceiving a slight delay compared to operating on-site, but commented that there was no particular discomfort in the controls. Figure 15 shows the construction site and the remote operation site.

5 Conclusions

This paper describes a method for transmitting multiple camera videos and remotely operating multiple construction machines in environments with limited network bandwidth and packet loss. As core technologies, we propose two techniques (BAERT): video coding technology for video transmission over limited bandwidth and error tolerance against transmission errors.

In the proof-of-concept experiments, we transmitted video from a total of 10 cameras—eight HD cameras mounted on two construction machines and two HD overhead cameras—to a remote location 900 km away via a private 5G network and Starlink Mini installed at the construction site. We confirmed that high-quality video without degradation could be transmitted with end-to-end latency below 170 ms, enabling efficient remote operation of the construction machinery.

Moving forward, we aim to further streamline efficiency from installation to operation on-site by automating various settings and configurations without any professional video transmission skills.

Acknowledgments

This study was carried out through the joint research work with Public Works Research Institute, Jitsuta Chugoku Co. Ltd., and Chudenko Corporation on productivity improvement of mechanical earthwork using OPERA (Open Platform for Earthwork with Robotics and Autonomy). The authors would like to thank all for discussions and support of the experiments.

This research was supported by the Ministry of Land, Infrastructure, Transport and Tourism's SBIR (Small Business Innovation Research) Construction Technology Research and Development Grant Program (JPJ000094).

References

[1] Japanese Infrastructure Ministry Unveils Action Plan “ i-Construction 2.0 ” to Achieve Construction Site Automation by 2040. On-line: https://www.decn.co.jp/?p=163015, Accessed: 01/12/2025.

[2] S. Yamaguchi, et.al. Direct-Visual-Operation Support System for Unmanned Construction. *The Open Civil Engineering Journal*, Vol.14, pages 105-112, 2020.

[3] K. Chayama, et.al. Technology of Unmanned Construction System in Japan. *Journal of Robotics and Mechanics*, Vo.26 NO.4: 403-417, 2014.

[4] Y. Nakajima, et.al. Ultra Low Delay and Resilient Multi-channel Video Encoder for Efficient Remote Construction. *ISARC 2025*, pages 516-523, 2025.

[5] S.R.Ely. MPEG video coding A simple introduction, EBU Technical Review Winter 1995, https://tech.ebu.ch/docs/techreview/trev_266-ely.pdf, Accessed: 01/12/2025.

[6] L. Landon, et.al. Enhancing 5G Performance: Reducing Service Time and Research Directions for 6G Standards. *6GNet*, pages 37-45, 2024.

[7] S.Kang, ey.al. Impact of FEC overhead on scalable video streaming. *NOSSDAV*, pages 123–128, 2005.

[8] G. Yamauchi, et.al Proposal of an Open Platform for Autonomous Construction Machinery Development. *Proceedings of the 40th ISARC*, pages 186-191, 2023.

[9] Hytec Inter Private 5G. https://hytecinter.com, Accessed: 01/12/2025.

[10] Starlink Mini Specifications. https://www.starlink.com, Accessed: 01/12/2025.

[11] S. Ito, et.al. Evaluation of work efficiency in the unmanned construction with communication delays. *The 16th Symposium on Construction Robotics in Japan*，No.06-1，2016.