Dynamic Masking and Unmasking for Drone Surveillance: A Privacy-Preserving Solution for Real-Time Video Processing

Yergali Berdibayev
AI & BigData Research Center
Gaion Co., Ltd.
Daejeon, Republic of Korea
ORCID: 0000-0001-7458-7009

Duc Tiep Vu
AI & BigData Research Center
Gaion Co., Ltd.
Daejeon, Republic of Korea
vuductiep@gaion.kr

Heon Gyu Lee[1]
AI & BigData Research Center
Gaion Co., Ltd.
Daejeon, Republic of Korea
hglee@gaion.kr

Abstract. **The proliferation of drone technology in surveillance, media, and commercial applications has intensified the need for robust privacy protection measures, especially in regions with strict data protection laws like the Republic of Korea. This paper introduces a versatile video masking and unmasking system designed for real-time processing of drone footage, capable of detecting and masking sensitive objects such as faces and vehicles during flight or live video streams. Leveraging state-of-the-art object detection algorithms, including YOLOv8, the system automatically identifies these objects and applies various masking techniques to obscure them, ensuring compliance with privacy regulations. Additionally, the system includes secure unmasking functionality for authorized users, enabling controlled access to unaltered footage when necessary. The effectiveness and efficiency of the system are demonstrated through various real-world scenarios, highlighting its adaptability to different environments and its potential applications in public safety, media, and commercial surveillance. The paper also discusses future directions for enhancing the system's capabilities and expanding its use cases to further advance privacy-preserving solutions and optimize performance.**

Keywords: computer vision, dynamic masking, drone surveillance, YOLOv8, RLE.

I. INTRODUCTION

The increasing use of drones in sectors such as surveillance, media, and commercial applications has raised significant concerns regarding privacy, especially in jurisdictions with stringent data protection regulations like the Republic of Korea [1]. While drone technology offers substantial benefits in various fields, its potential to infringe on individual privacy necessitates the development of advanced privacy protection mechanisms [2].

Traditional methods of de-identifying sensitive information often struggle to meet the real-time demands posed by live drone footage, where rapid detection and masking of personal data, such as faces and vehicles, are critical for compliance with privacy laws [3]. This paper addresses these challenges by introducing a sophisticated video masking and unmasking system tailored for real-time drone video processing. By automatically detecting and masking sensitive objects in live streams, this system aims to ensure legal compliance while balancing the need for operational efficiency and accuracy. The innovation of this work lies in its ability to provide robust privacy protection while offering authorized users the capability to securely restore unaltered footage when required.

In addition to the core functionalities, this paper extends the analysis of system performance through comparisons with other state-of-the-art object detection models and examines legal and ethical implications, extended applications, and potential optimizations for edge computing and security.

II. SYSTEM ARCHITECTURE

The proposed system consists of two main components: the masking system, which integrates object detection and masking processes, and the unmasking system, designed for efficient real-time operation.

The masking system begins by analyzing the input video stream from a drone using advanced object detection algorithms like YOLOv8, which excels in identifying specific objects such as faces, vehicles, and other privacy-sensitive elements. Once detected, a pixel-based mask is dynamically applied to the identified regions within each video frame, obscuring personal information to meet legal privacy requirements [4].

To enhance the system's performance, the architecture incorporates a compressed mask map that highlights the regions requiring de-identification. Instead of transferring a full-resolution mask for each frame, the system employs Run-Length Encoding (RLE) to compress the mask map into a more compact descriptor [5]. This approach significantly reduces bandwidth and storage requirements while preserving the system's capability to process data swiftly and efficiently [6]. For authorized personnel, the unmasking system decodes the RLE-compressed mask map and reconstructs the original image by restoring the masked regions, enabling selective access to unaltered footage where legally permissible.

As shown in Figure 1, our system architecture facilitates seamless interaction between multiple components. The drone captures video data and streams it via LTE/4G to the AI server using RTSP protocols. The AI server, equipped with object detection and masking models, processes the video streams to obscure sensitive information dynamically. The processed videos are then stored, either masked or unmasked, in a secure storage system. For standard users, a masked stream is delivered through a web-based real-time communication (RTC) interface, ensuring privacy

[1] Corresponding author

compliance. In contrast, authorized admins can access unmasked streams through the AI Drone Controller System, enabling them to review original footage when necessary. This architecture ensures that video data is managed efficiently and securely, with clear delineation between masked and unmasked access.

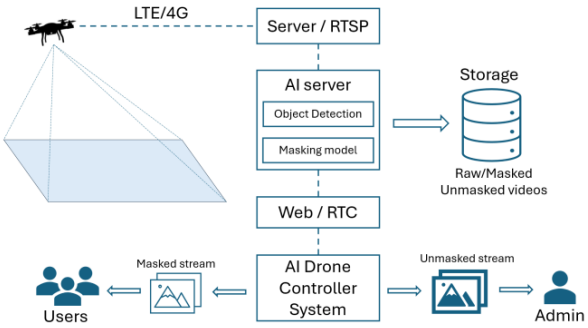


Fig. 1. System Architecture Diagram

Figure 1 visually encapsulates this end-to-end workflow, emphasizing the robust interplay between drone data capture, AI-driven processing, and secure video management, as well as the distinct pathways for user and admin access.

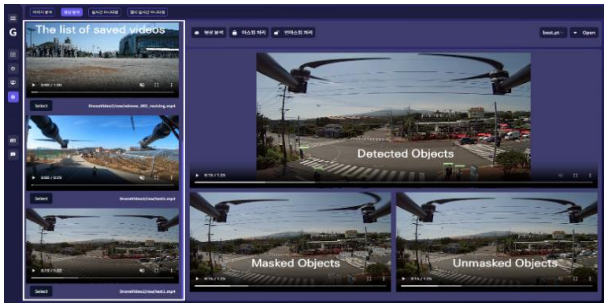


Fig. 2. AI Drone Controller System

To further illustrate the operational capabilities of the proposed system, Figure 2 shows the AI Drone Controller System's user interface. This interface highlights how the system manages real-time video streams by displaying detected objects, masked objects for privacy compliance, and unmasked objects for authorized viewing. The layout includes a sidebar with a list of saved videos, demonstrating the system's functionality for video management. This visualization underscores the system's core functions of object detection, masking, and unmasking, providing a practical view of how the architecture is implemented in a real-world application.

III. Methodology

The methodology involves three core processes: object detection, masking, and unmasking.

- *Object Detection:* The system leverages YOLOv8 for high-speed, high-accuracy detection of privacy-sensitive objects in live video streams, ensuring effective identification of privacy risks as they occur [7].
- *Masking:* Detected sensitive regions are masked using a binary mask map, where sensitive pixels are marked and compressed using RLE [8]. This process drastically reduces the size of the mask map without loss of critical information. The compressed mask map is transmitted alongside the encoded video stream, effectively obscuring sensitive areas.
- *Unmasking:* For authorized users, the unmasking system decodes the RLE-compressed mask map to identify and restore masked pixels using the original video and a securely stored key image, ensuring only authorized access to unaltered footage.

To demonstrate how the system operates in real-world scenarios, we present the following example:

Imagine a drone monitoring an intersection where a traffic accident occurs (as depicted in Figure 3). The drone captures the entire scene, including a witness car at the intersection. To protect the privacy of the individuals inside the witness car, the drone's system immediately detects and identifies the car, its occupants, and the license plate as privacy-sensitive elements. These identified areas are automatically masked in real-time to comply with privacy regulations.

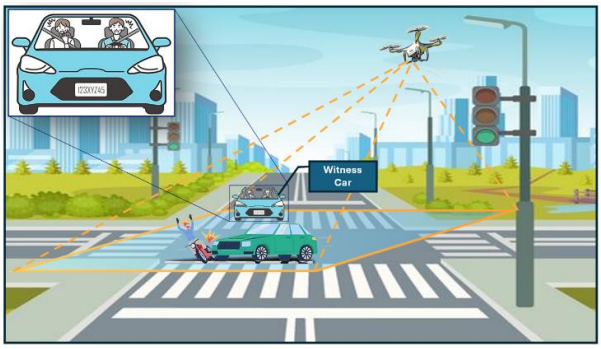


Fig. 3. Drone Surveillance Capturing a Traffic Accident and Privacy-Sensitive Objects (Witness Car)

Figure 3 illustrates the scenario of the drone capturing footage of the witness car at an accident scene. The drone identifies privacy-sensitive objects such as the faces of the occupants and the vehicle's license plate. The system ensures that these elements are obscured during real-time recording, thus maintaining the privacy of bystanders while still allowing comprehensive coverage of the event.

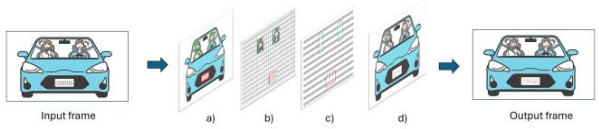


Fig. 4. Step-by-Step Process of Detecting and Masking Privacy-Sensitive Information in Video Frames

Figure 4 provides a detailed step-by-step process showing how the system detects and masks sensitive data in the captured video frames:

a) The initial input frame shows the detected sensitive areas, including the two faces of the car occupants and the license plate, highlighted by bounding boxes.

b) The system generates a binary mask map for the frame, where sensitive areas are identified by pixels marked as 1, and all non-sensitive areas are marked as 0. This mask map is presented as a grid, which delineates the sections of the frame that require obscuring to ensure privacy protection [9].
c) The mask map is compressed using Run-Length Encoding (RLE), which reduces the amount of data needed to represent the masked regions. This compression step is critical for maintaining system performance, especially in real-time video processing environments.
d) The final output frame is produced, in which the sensitive areas (the faces and the license plate) are successfully masked. This output ensures that privacy is preserved without compromising the visual clarity of non-sensitive parts of the footage.

IV. SYSTEM PERFORMANCE ANALYSIS

A. Running Environment and Dataset

The system was implemented and tested in a high-performance computing environment equipped with an Intel(R) Xeon(R) Gold 6248R CPU running at 3.00 GHz, an Nvidia A5000 GPU, and 128GB of RAM, selected to handle real-time processing demands. The evaluation used a diverse real-world dataset comprising drone footage from urban and rural settings, featuring various objects such as faces, vehicles, and other privacy-sensitive elements. The dataset was compiled from multiple sources, including publicly available datasets such as COCO and Roboflow, along with our own collected drone footage, resulting in a total of nearly 10,000 images to ensure comprehensive coverage of different environments. We utilized our own CVAT server (Computer Vision Annotation Tool) to manually annotate the collected footage, creating bounding boxes and labels in YOLO format, which were used for training and testing purposes. The dataset was split into 80% for training and 20% for testing, allowing robust evaluation and comparative analysis between YOLOv7 and YOLOv8.

B. Comparative Performance Analysis

1) Object Detection Models: YOLOv7 vs. YOLOv8

Performance was assessed using mean Average Precision (mAP) metrics at different Intersection over Union (IoU) thresholds. YOLOv8 generally outperformed YOLOv7, especially in person detection, with mAP50 scores of 0.719 compared to 0.620 for YOLOv7. However, YOLOv8 showed slightly reduced performance in vehicle detection.

TABLE I. OBJECT DETECTION MODEL PERFORMANCE

Objects	YOLOv7		YOLOv8	
	mAP_{50}	mAP_{50-95}	mAP_{50}	mAP_{50-95}
Total	0.766	0.599	0.796	0.617
Person	0.620	0.466	0.719	0.533
Car	0.912	0.733	0.873	0.700

2) Latency and Resource Consumption

In a real-time environment, reducing latency is crucial [10]. The introduction of RLE-based compression significantly reduces the bandwidth required to transmit the mask data alongside the video stream. By compressing consecutive sequences of identical pixel values in the mask map, RLE achieves high compression ratios with minimal computational overhead, enabling the system to maintain low-latency processing even at higher resolutions and frame rates.

Testing in a high-performance computing environment equipped with an Nvidia A5000 GPU and an Intel Xeon CPU demonstrated the system's ability to maintain real-time processing across various operational scenarios. YOLOv8 maintained an average latency of 50 milliseconds per frame at 1080p resolution, which is acceptable for real-time processing. GPU resources were efficiently utilized, averaging 60% usage during peak processing, while offloading most computational load from the CPU to ensure consistent performance.

TABLE II. SYSTEM PERFORMANCE WITH AND WITHOUT RLE-BASED COMPRESSION

Metric	With RLE	Without RLE
Latency	50 ms	Higher
Bandwidth Consumption	Significantly reduced	Higher
Computational Overhead	Minimal	Higher
GPU Utilization	60%	Higher (due to increased computational load)
CPU Offloading	Significant	Lower
Overall Performance	Real-time processing maintained	Potential degradation in real-time performance

3) Edge Cases and Failure Analysis

Edge cases such as poor lighting, high object density, occlusions, and overlapping objects were identified as challenges. The system's performance degraded under low-light conditions, with detection accuracy dropping by approximately 15% for faces and 10% for vehicles. Current efforts are focused on integrating pre-processing techniques and enhancing detection models to better handle these conditions, including multi-frame analysis to improve consistency.

TABLE III. EDGE CASES AND PERFORMANCE IMPACT IN OBJECT DETECTION

Edge Case	Challenge	Impact on Performance	Potential Improvements
Poor Lighting	Reduced visibility, especially at night or in shadows	Detection accuracy drops by 15% for faces and 10% for vehicles	Integration of pre-processing techniques (e.g., noise reduction, histogram equalization), use of IR cameras, or multi-frame analysis
High Object Density	Overcrowded scenes with multiple overlapping objects	Increased false positives and missed detections	Enhanced object detection algorithms (e.g., multi-scale detection) and post-processing refinement
Occlusions	Partial obstructions of objects (e.g., faces behind objects)	Reduced detection accuracy due to hidden features	Multi-frame analysis, leveraging temporal consistency, and object tracking
Overlapping Objects	Multiple objects of the same type appearing close to each other	Ambiguous detections leading to inaccurate bounding boxes	Refinement of bounding box merging techniques and using depth estimation for separation

Low Contrast Scenes	Scenes where objects blend with the background	Difficulty in distinguishing objects from the environment	Applying contrast enhancement and edge detection techniques to improve object visibility

4) Comparison with Other Privacy-Preserving Techniques

The system was benchmarked against other privacy-preserving methods, including blurring and pixelation [11], which often leave identifiable features visible. Our system demonstrated superior accuracy and processing speed, making it more suitable for real-time compliance needs.

C. Scalability and Edge Computing Potential

The system's architecture supports edge computing, allowing deployment on lower-power devices like the Nvidia Jetson series, suitable for real-world drone applications requiring edge processing. Integration with 5G/6G technologies is also being explored to further reduce latency and enhance real-time capabilities, making the system adaptable across different operational contexts.

D. Future Optimization Plans

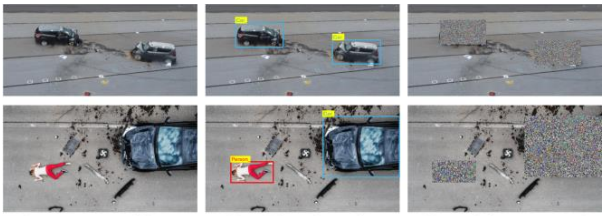
Future work includes optimizing neural networks for lower latency and resource consumption through pruning and quantization, enhancing adaptability across various hardware configurations to ensure effective privacy-preserving video processing.

V. Legal and Ethical Discussion

The system aligns with privacy laws across jurisdictions, including GDPR [12], CCPA [13], and local regulations in Korea [14, 15], by ensuring real-time masking of sensitive data and restricting access to unaltered footage. Ethical considerations include the risks of misuse and masking errors. To mitigate these, the system incorporates strict access controls, transparency measures, and error-checking protocols.

VI. Extended applications and Use Cases

Beyond surveillance, the system has applications in smart cities [16], autonomous vehicles [17], healthcare [18], and digital finance [19]. It enables privacy-preserving image and video processing that enhances safety and compliance without compromising individual privacy. Additionally, its post-processing capabilities in media production offer opportunities for selective obfuscation in public broadcasts.



a) Drone surveillance b) Object detection b) Masking approach

Fig. 5. Use Case Example: Drone surveillance of a car accident

Figure 5 illustrates a practical use case of the proposed system in drone surveillance during a car accident scenario. The figure demonstrates the sequence of operations:

a) *Drone Surveillance:* The initial raw footage captured by the drone.
b) *Object Detection:* The system automatically detects privacy-sensitive objects such as vehicles and individuals involved in accidents.
c) *Masking Approach:* Identified objects are masked to protect privacy, ensuring compliance with data protection regulations while maintaining the operational utility of the surveillance footage.

This example highlights the effectiveness of the system in real-world applications, where privacy-sensitive information is dynamically identified and appropriately masked in real-time drone video streams.

VII. System optimization and Scalability

Future optimizations aim to deploy on low-power, edge-based devices, leveraging 5G/6G technologies to enhance transmission speeds and reduce latency, making real-time privacy protection feasible even on constrained hardware. Research into adaptive models will enable the system to dynamically adjust to environmental conditions, regulatory requirements, or user preferences.

VIII. Security and Integrity measures

To secure unmasked footage, the system employs advanced encryption and secure access protocols, ensuring only authorized users can access or restore original content. Tamper detection mechanisms verify video integrity, maintaining authenticity and reliability for legal and operational purposes [20].

IX. Future directions

Further developments will explore AI-driven masking criteria based on real-time risk assessments and advanced anonymization techniques, such as recognizing individuals by clothing or gait [21]. Enhancements in environmental context awareness will improve the system's adaptability to factors like weather and time of day.

X. Interdisciplinary integration

Human-Computer Interaction (HCI) research is guiding the refinement of the system's interface, ensuring usability for non-expert drone operators. Additionally, considerations of broader social implications, including public perception and policy influence, are integral as privacy-preserving technologies evolve alongside societal views.

Acknowledgment

This work was supported by "Development of drone-robot cooperative multimodal delivery technology for cargo with a maximum weight of 40kg in urban areas" project of the Korea AeroSpace Administration (Project No, 00256794).

References

[1] Ko, H., Leitner, J., Kim, E., & Jeong, J. (2017). Structure and enforcement of data privacy law in South Korea. *International Data Privacy Law*, *7*(2), 100-114.

[2] R. Majeed, N. A. Abdullah, M. F. Mushtaq and R. Kazmi, "Drone Security: Issues and Challenges." Parameters, 2(5) 2021.

[3] Erkin, Z., Franz, M., Guajardo, J., Katzenbeisser, S., Lagendijk, I., & Toft, T. (2009). Privacy-preserving face recognition. In *Privacy Enhancing Technologies: 9th International Symposium, PETS 2009,*

Seattle, WA, USA, August 5-7, 2009. Proceedings 9 (pp. 235-253). Springer Berlin Heidelberg.

[4] Liu, L., Yu, E., & Mylopoulos, J. (2003, September). Security and privacy requirements analysis within a social setting. In *Proceedings. 11th IEEE International Requirements Engineering Conference, 2003.* (pp. 151-161). IEEE.

[5] VidyaSagar, M., & Victor, J. R. (2013). Modified run length encoding scheme for high data compression rate. *Int J Adv Res Comput Eng Technol (IJARCET)*, *2*, 12.

[6] A. Birajdar, H. Agarwal, M. Bolia and V. Gupte, "Image compression using run length encoding and its optimization," IEEE Global Conference for Advancement in Technology (GCAT), pp. 1-6, October 2019. DOI:10.1109/GCAT47503.2019.8978464

[7] G. Wang, Y. Chen, P. An, H. Hong, J. Hu and T. Huang, "UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios". Sensors, 23(16), 2023. DOI:10.3390/s23167190

[8] A. R. Idrais, I. Aljarrah, and O. Al-Khaleel, "A spatial image compression algorithm based on run length encoding." Bulletin of Electrical Engineering and Informatics 10(5), 2021.

[9] M. Iliadis, L. Spinoulas and A. K. Katsaggelos, "Deepbinarymask: Learning a binary mask for video compressive sensing." Digital Signal Processing, 96, 102591, 2020.

[10] Mahmud, R., Ramamohanarao, K., & Buyya, R. (2018). Latency-aware application module management for fog computing environments. *ACM Transactions on Internet Technology (TOIT)*, *19*(1), 1-21.

[11] Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(1), 101-116.

[12] Voigt, P., & Von dem Bussche, A. (2017). The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, *10*(3152676), 10-5555.

[13] Goldman, E. (2020). An introduction to the california consumer privacy act (ccpa). *Santa Clara Univ. Legal Studies Research Paper*.

[14] Park, S., Choi, G. J., & Ko, H. (2020). Information technology–based tracing strategy in response to COVID-19 in South Korea—privacy controversies. *Jama*, *323*(21), 2129-2130.

[15] Berdibayev, Y., & Kwon, Y. (2021). Fear of COVID-19, social isolation and digital financial services during the COVID-19 pandemic: The unified theory of acceptance and use technology (UTAUT) model.

[16] Van Zoonen, L. (2016). Privacy concerns in smart cities. *Government Information Quarterly*, *33*(3), 472-480.

[17] Hataba, M., Sherif, A., Mahmoud, M., Abdallah, M., & Alasmary, W. (2022). Security and privacy issues in autonomous vehicles: A layer-based survey. *IEEE Open Journal of the Communications Society*, *3*, 811-829.

[18] Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. *Journal of big data*, *5*(1), 1-18.

[19] Berdibayev, Y., & Kwon, Y. (2020). Improving digital financial services inclusion: A panel data analysis.

[20] MacNeil, H. (2000). Providing grounds for trust: developing conceptual requirements for the long-term preservation of authentic electronic records. *Archivaria*, 52-78.

[21] Little, J., & Boyd, J. (1998). Recognizing people by their gait: the shape of motion. *Videre: Journal of computer vision research*, *1*(2), 1-32.