Architecture and Design Study with Analytical Validation of a Mixed Reality Software System for Pilot Training with Real-Time Visual Scene Substitution

Serhii D. Prykhodchenko
Department of Computer systems' software, Dnipro University of Technology, Ukraine

Oksana Yu. Prykhodchenko
Department of Economics and Economic Cybernetics, Dnipro University of Technology, Ukraine

Andrii A. Martynenko
Department of Computer systems' software, Dnipro University of Technology, Ukraine

Dmytro Babets
Department of Applied Mathematics, Dnipro University of Technology, Ukraine

Marcin Paszkuta
merQlab, Department of Computer Graphics, Vision and Digital Systems,
Silesian University of Technology, Gliwice, Poland

Dariusz Myszor
Department of Algorithmics and Software, Faculty of Automatic Control,
Electronics and Computer Science, Silesian University of Technology, Gliwice, Poland

Przemysław Skurowski
merQlab, Department of Computer Graphics, Vision and Digital Systems,
Silesian University of Technology, Gliwice, Poland

Krzysztof Cyran
merQlab, Department of Computer Graphics, Vision and Digital Systems,
Silesian University of Technology, Gliwice, Poland

Abstract

This paper presents the design and architectural formalization of a mixed reality (MR) system for pilot training that supports selective, region-level visual scene substitution. The proposed approach enables context-aware replacement of semantically meaningful regions within the pilot's egocentric view – such as cockpit displays and external windows – while preserving direct physical interaction with real cockpit controls and the pilot's hands.

The core contribution is a modular real-time architecture that integrates semantic perception, structured scene representation, and GPU-accelerated mask-based compositing into a unified dataflow pipeline. The system is designed with explicit latency-aware modelling to support analytical evaluation of performance under strict real-time constraints (targeting below 20–30 ms).

A functional prototype was implemented and evaluated under different hardware configurations. Experimental results demonstrate that the GPU-accelerated version running on an NVIDIA RTX 3060 achieves an average end-to-end latency of 26.3 ms (approximately 38 fps) with improved temporal stability. These results confirm the practical feasibility of the proposed architecture under controlled conditions.

Adopting a design-science perspective, this work combines analytical modelling, architectural design, and preliminary prototype validation. The study addresses the need for stronger technical grounding in XR-based aviation training research prior to large-scale human-subject evaluation. The presented approach provides a reproducible foundation for the development of hybrid training systems that combine the sensorimotor fidelity of physical cockpits with the instructional flexibility of real-time visual substitution.

© Serhii D. Prykhodchenko, Oksana Yu. Prykhodchenko, Andrii A. Martynenko, Dmytro Babets, Marcin Paszkuta, Dariusz Myszor, Przemysław Skurowski, and Krzysztof Cyran;
licensed under Creative Commons License CC-BY 4.0
7th International Computer Programming Education Conference (ICPEC 2026).
Editors: Filipe Portela, Luís Matos, and Tiago Guimarães; Article No. 15; pp. 15:1–15:11
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

2012 ACM Subject Classification Computer systems organization → Real-time systems; Human-centered computing → Mixed / augmented reality; Computing methodologies → Computer vision; Computing methodologies → Image segmentation; Computer systems organization → Embedded and cyber-physical systems; Software and its engineering → Data flow architectures

Keywords and phrases Mixed Reality, Pilot Training, Visual Scene Substitution, Region-level Compositing, Real-time Architecture

Digital Object Identifier 10.4230/OASIcs.ICPEC.2026.15

Supplementary Material *Software*: `https://github.com/prykhodchenkosd/MRPilotRealtimeTest` archived at `swh:1:dir:c6e42056d4c3aa7bb17f93baecb3230bb20d1110`

1 Introduction

The training of commercial pilots requires a high degree of perceptual fidelity, sensorimotor realism, and the ability to safely rehearse a wide range of operational scenarios, including rare and safety-critical events. Traditional full-flight simulators provide high realism but are expensive, immobile, and difficult to reconfigure quickly. Virtual reality (VR) systems offer excellent scenario control, yet they typically eliminate direct tactile interaction with physical cockpit controls, which is important for developing accurate procedural memory [4, 10, 6]. Augmented reality (AR) approaches preserve the real environment but are generally limited to additive overlays and cannot meaningfully modify existing visual elements or instrument behaviour [2, 13].

Mixed reality (MR) systems have emerged as a promising direction by combining physical cockpit interaction with virtual content. However, most current MR implementations focus primarily on replacing the external visual scene visible through the cockpit windows, while leaving internal instrumentation largely unchanged [7, 12, 20, 5]. Although video see-through and passthrough-based systems provide greater potential for visual manipulation, they have rarely been used for structured, region-selective substitution of semantically meaningful areas inside the cockpit.

This work addresses the above limitations by proposing a modular real-time architecture for selective, region-level visual scene substitution. The approach aims to enable context-aware replacement of specific regions within the pilot's egocentric view – such as cockpit displays and external windows – while maintaining full physical interaction with the real controls and the pilot's hands.

To clarify the positioning of the proposed system, Table 1 compares the main immersive training paradigms.

Table 1 Structural comparison of immersive training paradigms for pilot training.

System Type	Physical Cockpit	Scene Modification	Region-Level Control	Dynamic Reconfiguration
VR	✗	Full	✗	✓
AR	✓	Additive only	✗	Limited
MR (classic)	✓	External only	✗	Limited
Proposed	✓	**Selective**	✓	✓

The problem addressed in this work can be formalised as the real-time synthesis of a perceptually coherent mixed visual stream under strict latency and sensorimotor constraints. At each discrete time step t, the system receives a multimodal input tuple (I_t, P_t, G_t), where

$I_t \in \mathbb{R}^{H\times W\times 3}$ denotes the captured egocentric RGB image, $P_t \in SE(3)$ the six-degree-of-freedom head pose, and $G_t \in \mathbb{R}^3$ the gaze direction. The objective is to produce an output image

$$O_t = f(I_t, P_t, G_t, \theta_t),$$

where θ_t denotes adaptive control parameters. The transformation operates over a dynamically inferred set of semantically meaningful regions with associated masks $M_t^k(x) \in [0,1]$, such that the composited output satisfies

$$O_t(x) = (1 - M_t(x))I_t(x) + M_t(x)V_t(x),$$

with $M_t(x) = \sum_k M_t^k(x)$ (normalized to $[0,1]$) and V_t the rendered virtual content. The system is designed to satisfy end-to-end latency $T_{\text{total}} \leq T_{\text{frame}}$ (targeting below 20–30 ms) while ensuring temporal stability and geometric alignment between real and virtual elements.

The main contributions of this work are as follows:

- A formally defined approach to region-level visual substitution for mixed reality training systems, enabling selective replacement of semantically meaningful regions.
- A modular real-time architecture with explicit separation of perception, scene representation, and rendering layers, following a dataflow-driven design.
- A GPU-centric pipeline for low-latency semantic segmentation and mask-based compositing.
- An analytically grounded performance model for evaluating latency and concurrency in the pipeline.
- A prototype implementation and preliminary experimental evaluation under different hardware configurations.
- A design-science framework that combines analytical validation with a pathway toward future empirical studies.

This study aims to contribute a technically grounded foundation for hybrid mixed reality systems in pilot training.

2 Related Work

Immersive technologies have been increasingly explored to enhance pilot training, yet most solutions still fail to simultaneously provide high sensorimotor fidelity, fine-grained control over the visual environment, and real-time semantic adaptability. This section critically reviews Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) approaches, with particular emphasis on their limitations regarding region-level visual substitution inside a physical cockpit.

2.1 Virtual Reality in Pilot and Drone Training

VR-based simulators offer excellent scenario control and have shown benefits in procedural learning and cost reduction [6, 10, 4, 16, 18]. Systems have been developed for both manned aircraft [4, 16] and drone pilot training [1]. Systematic reviews confirm that while VR improves immersion and allows training of rare events, the complete detachment from physical cockpit hardware often weakens the transfer of procedural and muscle memory [18, 10]. In contrast to the proposed architecture, VR operates at a full-scene level and cannot preserve the tangible interaction with real controls and instruments.

2.2 Augmented Reality for Instructional Support

AR solutions enhance the real cockpit with overlaid digital information, demonstrating improvements in situational awareness, procedural performance, and cognitive load reduction [2, 13, 17]. However, because AR is limited to additive overlays, it cannot replace or modify existing visual elements such as instrument displays or the external view through windows. This fundamental constraint restricts its use for dynamic failure injection and realistic scenario adaptation [13, 11]. The proposed region-level substitution paradigm overcomes this limitation by enabling replacement rather than mere augmentation of semantically meaningful regions.

2.3 Mixed Reality and Video See-Through Systems

Mixed Reality systems that integrate physical cockpits with virtual content represent the most relevant prior work. Several recent implementations replace the external visual scene while preserving physical controls [7, 12, 20, 21]. Ernst et al. [7] developed a video see-through MR flight simulator, while Kimura et al. [12] and Zintl et al. [20, 21] focused on instructor operator stations and motion-based MR simulators for advanced air mobility (eVTOL). Other studies explored MR applications in general aviation [11] and multi-user training scenarios [9].

Despite these contributions, the majority of existing MR systems remain limited in scope: they primarily substitute only the external view and leave cockpit instrumentation unchanged. Video see-through approaches [7, 19] enable deeper visual manipulation but are rarely combined with semantic understanding for selective, context-aware region substitution. Most implementations lack a modular architecture that explicitly separates perception, structured scene representation, and GPU-accelerated compositing. Consequently, instructors cannot dynamically modify display content or inject instrument failures without physical modifications to the cockpit.

Human factors research further highlights challenges such as latency, registration errors, cybersickness, and limited instructor control in current MR aviation simulators [8, 5, 14, 15]. Systematic reviews of XR in pilot training [4, 18, 16] consistently emphasise the need for better integration of real and virtual elements while maintaining physical fidelity – a gap the present work directly addresses.

2.4 Positioning of the Proposed Architecture

The proposed system advances beyond prior work by introducing a dedicated real-time architecture for *semantically grounded, region-level visual substitution.* Unlike full-scene VR solutions [6, 18], additive AR systems [2, 13], and most existing MR implementations that focus on external view replacement [7, 20, 12], this architecture enables selective replacement of both external windows and internal cockpit displays while preserving full physical interaction with controls and the pilot's hands.

Unlike prior VR-, AR-, and MR-based training systems, the proposed architecture combines semantically controlled region-level substitution, manipulation of both external and internal cockpit elements, and an explicit latency-aware analytical execution model within a unified real-time pipeline. Existing MR approaches primarily focus on replacement of the external visual scene, while cockpit instrumentation typically remains physically unchanged and semantically unmanaged.

By combining real-time semantic segmentation, structured scene representation, GPU-accelerated mask-based compositing, and an explicit latency-aware dataflow model, the presented design offers greater instructional flexibility and finer visual control than previous

Table 2 Comparison of the proposed architecture with representative immersive pilot training systems.

Work	External View	Internal Displays	Semantic Regions	Real-time	Analytical Model
VR simulators [4, 16, 1, 6]	✓	✗	✗	✓	✗
AR cockpit systems [2, 13, 19, 17]	✓	✓	✗	✓	✗
MR view substitution & blending [7, 12, 20, 21]	✓	✓	✗partial	✓	✗partial
General MR/AR aviation training [11, 9]	✓	✓partial	✗	✓	✗
Proposed architecture	✓	✓	✓	✓	✓

systems. Furthermore, the integration of analytical performance modelling with prototype validation distinguishes this contribution from predominantly conceptual or purely empirical studies in the field [5, 8, 3].

3 System Architecture

The proposed system implements a real-time mixed reality pipeline for selective region-level visual substitution in a pilot training environment. The architecture is designed as a modular dataflow system that jointly processes visual perception, scene representation, and GPU-based rendering under strict latency constraints.

Let the input at time t be defined as a multimodal tuple:

$$F_t = (I_t, P_t, G_t), \tag{1}$$

where I_t is the captured RGB frame, $P_t \in SE(3)$ is the head pose, and G_t is the gaze vector.

The system is formulated as a sequential dataflow pipeline:

$$F_t \to C_t \to M_t \to S_t \to \tilde{S}_t \to V_t \to O_t, \tag{2}$$

where C_t denotes capture, M_t semantic perception output, S_t structured scene representation, $\tilde{S}_t$ substitution-aware representation, V_t rendered virtual content, and O_t the final composited output. Compositing is included within the output stage O_t.

We distinguish between two latency models: the ideal sequential latency and the execution-aware latency under pipeline parallelism.

The ideal end-to-end latency is defined as:

$$T_{total} = T_{cap} + T_{seg} + T_{scene} + T_{subst} + T_{rend} + T_{disp}. \tag{3}$$

Here:
- T_{cap} – image acquisition latency,
- T_{seg} – semantic segmentation latency,
- T_{scene} – structured scene construction,
- T_{subst} – substitution reasoning and mask fusion,
- T_{rend} – GPU rendering of virtual content,
- T_{disp} – compositing, scan-out, and XR display synchronization.

Due to asynchronous execution and GPU/CPU overlap, the effective latency is defined as:

$$T_{eff} = \max(T_{seg}, T_{rend}) + T_{subst} + T_{disp}. \tag{4}$$

This formulation reflects the critical-path behavior of the system under parallel execution.

Real-time constraints require:

$$T_{eff} \leq 33\,ms. \tag{5}$$

Temporal stability is evaluated over a sliding window of N frames as:

$$\sigma_T = \mathrm{Var}(\{T_{total}^{(i)}\}_{i=1}^{N}), \tag{6}$$

capturing system jitter over time.

The substitution module transforms semantic perception into render-ready control signals. It performs:

- semantic-to-render mapping,
- mask fusion and normalization,
- overlap resolution between regions,
- gaze- and context-aware weighting.

Semantic perception output is defined as a structured set of region masks:

$$M_t = \{R_t^1, R_t^2, \ldots, R_t^k\}, \tag{7}$$

where each R_t^i is a binary or probabilistic region mask corresponding to a detected semantic class or instance.

The substitution-aware representation is defined as a constructive mapping:

$$\tilde{S}_t = \mathcal{F}(M_t, S_t, G_t), \tag{8}$$

where $\mathcal{F}$ is a rule-based or learned fusion operator integrating spatial, semantic, and gaze-driven constraints.

The final compositing operation is defined as:

$$O_t(x) = (1 - \alpha_t(x))I_t(x) + \alpha_t(x)V_t(x), \tag{9}$$

with blending mask $\alpha_t(x)$ constrained to semantically valid regions:

$$\mathrm{supp}(\alpha_t) \subseteq \bigcup_i R_t^i. \tag{10}$$

The system operates under a hybrid execution model:

- GPU: segmentation, rendering, compositing,
- CPU: control logic and scene management,
- XR runtime: tracking and display synchronization.

The architecture enables pipeline overlap between perception and rendering stages, improving throughput under real-time constraints.

The architecture separates perception, representation, substitution, rendering, and compositing into modular layers, enabling real-time performance while supporting semantically controlled region-level visual manipulation of the egocentric field of view.

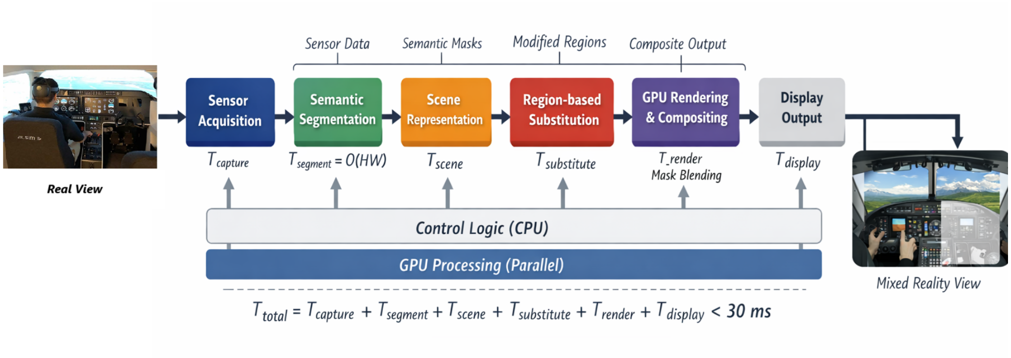

Figure 1 Latency-aware dataflow architecture of the proposed mixed reality substitution system.

4 Evaluation and Experimental Results

To address the limitations of purely conceptual validation frequently observed in mixed reality training research, the proposed architecture was evaluated using a reproducible real-time prototype implementation developed in Python with GPU-accelerated rendering and semantic segmentation. The complete benchmarking framework, including automated latency profiling and execution logging, is publicly available at `https://github.com/prykhodchenkosd/MRPilotRealtimeTest`. The evaluation focuses on validating the practical feasibility of region-level visual substitution under real-time constraints and on quantifying the impact of hardware acceleration on temporal stability and overall system responsiveness.

Three execution configurations were evaluated: a lightweight baseline pipeline without computationally intensive semantic inference, a CPU-based YOLO segmentation pipeline, and a GPU-accelerated YOLO pipeline executed on an NVIDIA RTX 3060 GPU. All configurations used the same rendering and compositing architecture to ensure direct comparability. Measurements were collected under continuous streaming conditions using frame-wise profiling of the capture, segmentation, rendering, and compositing stages. For each stage, both the mean execution time and the standard deviation were recorded to characterize not only throughput but also temporal stability, which is critical for perceptual consistency in immersive mixed reality systems.

The baseline configuration demonstrates the minimum processing overhead achievable by the proposed architecture under simplified semantic conditions. In this mode, the system achieves an average total latency of 27.92 ms and an average frame rate of 36.96 FPS. The segmentation stage contributes only 3.94 ms on average, while compositing remains the dominant computational stage due to image blending and synchronization overhead. Despite the absence of full semantic processing, this configuration establishes a lower-bound reference for the architectural pipeline itself and confirms that the rendering and compositing framework can operate within interactive real-time constraints.

The CPU-based YOLO configuration introduces substantially higher computational cost and temporal instability. Semantic segmentation becomes the dominant bottleneck, with an average latency of 30.43 ms and a large standard deviation of 48.29 ms. This variability propagates through the execution pipeline and results in a total average latency of 64.79 ms with a high temporal variance of 70.52 ms. The corresponding frame rate decreases to 25.36 FPS, approaching the lower acceptable boundary for interactive mixed reality applications. The obtained measurements indicate that CPU-only execution is insufficient for stable region-level visual substitution in semantically driven mixed reality scenarios, particularly under continuous streaming conditions where latency spikes negatively affect temporal coherence.

In contrast, the GPU-accelerated configuration executed on an NVIDIA RTX 3060 demonstrates a substantial reduction in both latency and temporal variance across all stages of the pipeline. Segmentation latency decreases to 6.85 ms with low variability, while GPU-based compositing reduces compositing overhead to 4.12 ms. The overall end-to-end latency decreases to 26.30 ms, satisfying the target real-time constraint defined for immersive mixed reality interaction. The achieved frame rate of 38.12 FPS confirms stable interactive operation, while the significantly reduced standard deviation of total latency (3.85 ms) indicates reduced latency variability and smoother temporal behavior. These results confirm that GPU acceleration not only improves throughput but also substantially stabilizes frame delivery, which is essential for perceptual coherence and user comfort in mixed reality environments.

Segmentation quality was evaluated through systematic visual inspection of the output frames. The system produced stable and well-defined region masks for both windows and cockpit displays under controlled conditions, enabling coherent real-time visual substitution. Although a formal quantitative assessment using annotated datasets (e.g., mean IoU across diverse lighting conditions) is planned for future iterations, the current perception pipeline already satisfies the architectural and performance requirements of the proposed mixed reality system.

Across all evaluated configurations, the results therefore validate the feasibility of the proposed architecture as a real-time mixed reality substitution system capable of semantically controlled region-level manipulation while preserving interactive responsiveness.

Table 3 Performance evaluation of the proposed mixed reality pipeline under different execution configurations.

Configuration	**Capture (ms)**	**Segmentation (ms)**	**Rendering (ms)**	**Compositing (ms)**	**Total (ms)**	**FPS**
Simple Test	10.22 ± 5.53	3.94 ± 1.05	3.33 ± 0.60	10.43 ± 0.99	27.92 ± 5.50	36.96 ± 6.09
YOLO on CPU	1.23 ± 1.70	30.43 ± 48.29	7.27 ± 1.02	17.52 ± 1.46	64.79 ± 70.52	15.4 ± 7.43
YOLO on GPU (RTX 3060)	1.21 ± 1.84	6.85 ± 1.12	6.42 ± 0.78	4.12 ± 0.65	26.30 ± 3.85	38.12 ± 4.71

5 Limitations and Discussion

Although the proposed architecture and prototype implementation demonstrate the feasibility of real-time region-level visual scene substitution, several limitations should be acknowledged. First, the current evaluation was conducted using a simplified desktop-based prototype with a standard RGB camera and static virtual content rather than within a realistic cockpit simulator or immersive head-mounted display environment. As a result, the reported performance metrics, including the average end-to-end latency of 26.3 ms on an NVIDIA RTX 3060 GPU, reflect controlled laboratory conditions and may not directly generalize to operational training scenarios involving rapid head motion, complex illumination conditions, and dynamic cockpit interactions.

Second, the perception pipeline currently relies on relatively lightweight colour-based and YOLO-based semantic segmentation models. While these models are sufficient for validating the architectural feasibility of real-time region-level substitution, their robustness under realistic cockpit conditions has not yet been systematically evaluated. In particular, illumination variability, reflections from cockpit surfaces, motion blur, and partial occlusions caused by the pilot's hands or body may affect segmentation stability and temporal consistency. Improving robustness and domain adaptation of semantic perception therefore remains an important direction for future research.

Third, the present study primarily focuses on architectural design, analytical modelling, and preliminary technical validation. The proposed system has not yet been evaluated through controlled user studies involving pilots or flight instructors. Consequently, the work does not currently provide empirical evidence regarding training effectiveness, skill transfer, cognitive workload, situational awareness, or instructor usability. These aspects require dedicated human-subject experiments and represent a critical next step toward practical deployment.

From a methodological perspective, pixel-level segmentation accuracy metrics such as Intersection over Union (IoU) are not reported in this study, as the current prototype is not built upon a fully annotated ground-truth dataset. Instead, the focus is placed on real-time system performance, architectural feasibility, and temporal characteristics of the pipeline. A dataset-driven evaluation with formally defined ground truth annotations is left for future work, where more comprehensive benchmarking of perception accuracy will be required.

The proposed approach also inherits several limitations commonly associated with video see-through mixed reality systems. These include reduced visual fidelity compared to optical see-through displays, increased computational and energy requirements, and perceptual issues such as the vergence–accommodation conflict. Furthermore, although the GPU-accelerated implementation achieves interactive real-time performance, maintaining consistently low latency on untethered or resource-constrained hardware platforms remains a significant engineering challenge.

Despite these limitations, the presented work establishes a reproducible and analytically grounded foundation for semantically controlled region-level substitution in mixed reality pilot training. The modular architecture, explicit latency-aware design, and publicly available prototype framework provide a basis for further experimental development and comparative evaluation. Future work will focus on integration with immersive head-mounted displays and realistic cockpit environments, development of more robust domain-adapted perception models, incorporation of advanced occlusion and depth handling, and comprehensive user studies aimed at assessing training transfer, usability, and instructional effectiveness.

6 Conclusion and Future Work

This paper presented the architectural design, formalization, and preliminary technical validation of a mixed reality system for pilot training based on selective region-level visual scene substitution. The proposed approach enables context-aware replacement of semantically meaningful regions, such as cockpit displays and external windows, while preserving direct interaction with physical cockpit controls and the pilot's hands.

A modular latency-aware architecture was developed that integrates semantic perception, structured scene representation, and GPU-accelerated mask-based compositing within a real-time dataflow pipeline. The work additionally introduced an analytical execution model describing the relationship between segmentation, rendering, compositing, and overall end-to-end latency under parallel execution constraints.

A functional prototype implementation was evaluated under multiple execution configurations. Experimental results indicate that GPU acceleration substantially improves both throughput and temporal stability compared to CPU-only execution. In the evaluated configuration using an NVIDIA RTX 3060 GPU, the system achieved an average end-to-end latency of 26.3 ms with an average frame rate of approximately 38 fps, satisfying the targeted real-time interaction requirements under controlled laboratory conditions.

The presented results should be interpreted as preliminary technical validation of the architectural feasibility of the proposed approach rather than a complete assessment of training effectiveness. The current prototype was evaluated using a desktop-based setup with a standard RGB camera and static virtual content. Formal user studies involving pilots or flight instructors, as well as quantitative perception benchmarks using annotated datasets, were outside the scope of the present work.

Nevertheless, the study demonstrates that semantically controlled region-level substitution can be implemented within interactive real-time constraints using a modular GPU-oriented mixed reality pipeline. The proposed architecture and publicly available prototype framework provide a reproducible technical basis for further research on hybrid mixed reality training systems.

Future work will focus on integration with immersive video see-through head-mounted displays and realistic cockpit mock-ups, improvement of segmentation robustness under challenging lighting and motion conditions, incorporation of depth-aware occlusion handling, and comprehensive experimental evaluation with pilots and instructors. Additional directions include quantitative segmentation benchmarking, comparative studies against existing XR training paradigms, and exploration of collaborative multi-user training scenarios.

References

1 G. Albeaino, R. Eiris, M. Gheisari, and R. Issa. Dronesim: a VR-based flight training simulator for drone-mediated building inspections. *Construction Innovation*, 2021. `doi:10.1108/ci-03-2021-0049`.

2 D. Arjoni, I. De Souza Rehder, J. Figueira, and E. Villani. Augmented reality for training formation flights: An analysis of human factors. *Heliyon*, 9, 2023. `doi:10.1016/j.heliyon.2023.e14181`.

3 R. Bödding, S. Schriek, and G. Maier. A systematic review and meta-analysis of mixed reality in vocational education and training: examining behavioral, cognitive, and affective training outcomes and possible moderators. *Virtual Reality*, 29, 2025. `doi:10.1007/s10055-025-01118-z`.

4 J. Cross, C. Boag-Hodgson, T. Ryley, T. Mavin, and L. Potter. Using extended reality in flight simulators: A literature review. *IEEE Transactions on Visualization and Computer Graphics*, 29:3961–3975, 2022. `doi:10.1109/tvcg.2022.3173921`.

5 J. Cross and T. Ryley. Assessing evidence-based training in a collaborative virtual reality flight simulator. *The Aeronautical Journal*, 2024. `doi:10.1017/aer.2024.113`.

6 P. Dymora, B. Kowal, M. Mazurek, and S. Romana. The effects of virtual reality technology application in the aircraft pilot training process. *IOP Conference Series: Materials Science and Engineering*, 1024, 2021. `doi:10.1088/1757-899x/1024/1/012099`.

7 J. Ernst, T. Laudien, and S. Schmerwitz. Implementation of a mixed-reality flight simulator: blending real and virtual with a video-see-through head-mounted display. In *Proceedings of SPIE*, volume 12538, pages 125380R–125380R–10, 2023. `doi:10.1117/12.2664848`.

8 S. Fussell, S. Rebensky, and S. Mcgee. Human factors challenges for extended reality aviation training simulation. In *AHFE International*, 2025. `doi:10.54941/ahfe1006386`.

9 Y. Go, H. Kang, J. Lee, M. Yu, and S. Choi. Multi-user drone flight training in mixed reality. *Electronics*, 2021. `doi:10.3390/electronics10202521`.

10 R. Guthridge and V. Clinton-Lisell. Evaluating the efficacy of virtual reality (vr) training devices for pilot training. *Journal of Aviation Technology and Engineering*, 2023. `doi:10.7771/2159-6670.1286`.

11 Christopher Katins, Sebastian S. Feger, and Thomas Kosch. Exploring mixed reality in general aviation to support pilot workload. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3544549.3585742.

12 S. Kimura, M. Zintl, C. Hammann, and F. Holzapfel. Simulator-based mixed reality evtol pilot training: The instructor operator station. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024. doi:10.1145/3613904.3642060.

13 B. Moesl, H. Schaffernak, W. Vorraber, R. Braunstingl, and I. Koglbauer. Performance, emotion, presence: Investigation of an augmented reality-supported concept for flight training. *Applied Sciences*, 2023. doi:10.3390/app132011346.

14 A. Perez, A. Singh, J. Mitchell, and P. Swadling. Strategies to manage human factors in mixed reality pilot training: A survey. arXiv preprint arXiv:2507.15526, 2025. doi:10.48550/arxiv.2507.15526.

15 S. Rizvi, U. Rehman, S. Cao, and B. Moncion. Exploring technology acceptance of flight simulation training devices and augmented reality in general aviation pilot training. *Scientific Reports*, 15, 2025. doi:10.1038/s41598-025-85448-7.

16 G. Ross and A. Gilbey. Extended reality (xr) flight simulators as an adjunct to traditional flight training methods: a scoping review. *CEAS Aeronautical Journal*, 14:799–815, 2023. doi:10.1007/s13272-023-00688-5.

17 H. Schaffernak, B. Moesl, W. Vorraber, M. Holy, E. Herzog, R. Novak, and I. Koglbauer. Novel mixed reality use cases for pilot training. *Education Sciences*, 2022. doi:10.3390/educsci12050345.

18 A. Somerville, K. Joiner, T. Lynar, and G. Wild. Applications of extended reality in pilot flight simulator training: a systematic review with meta-analysis. *Visual Computing for Industry, Biomedicine, and Art*, 8, 2025. doi:10.1186/s42492-025-00206-w.

19 J. Wallace, Z. Hu, and D. Carroll. Augmented reality for immersive and tactile flight simulation. *IEEE Aerospace and Electronic Systems Magazine*, 35:6–14, 2020. doi:10.1109/maes.2020.3002000.

20 M. Zintl, S. Kimura, and F. Holzapfel. A mixed reality research flight simulator for advanced air mobility vehicles. *AIAA AVIATION FORUM AND ASCEND 2024*, 2024. doi:10.2514/6.2024-4651.

21 M. Zintl, S. Kimura, and F. Holzapfel. Development and human-centered evaluation: A motion-based mixed-reality flight simulator. *Journal of Aerospace Information Systems*, 2026. doi:10.2514/1.i011738.