ICJK
Chinese Journal of Information Fusion
http://dx.doi.org/10.62762/CJIF.2025.472730

RESEARCH ARTICLE

Adaptive Field-of-View Restriction Based on Sigmoid and Bézier Smoothing Control Functions

Wenli Shang[1,*] and **Kazlova Alena**[1]

[1] Faculty of Radiophysics and Computer Technologies, Belarusian State University, Minsk 220030, Belarus

Abstract

Virtual Reality (VR) systems require extremely low latency and accurate user intent prediction to ensure immersion. However, traditional visual modulation techniques primarily rely on single-sensor inputs (e.g., IMU angular velocity), which suffer from inherent mechanical delays and fail to capture complex spatial-temporal dynamics, ultimately leading to severe visually induced motion sickness (VIMS). Existing methods lack effective fusion of multi-source heterogeneous data, resulting in delayed responses and abrupt visual transitions. To address this, this study proposes a novel spatio-temporal information fusion framework for adaptive visual modulation in VR. The proposed architecture deeply fuses high-frequency eye-tracking signals (physiological intent), IMU data (physical kinematics), and dense optic flow (environmental spatial features). Specifically, a cross-modal temporal intent prediction model is established to fuse gaze and head-motion data, achieving early action anticipation. Simultaneously, a spatial data fusion module dynamically allocates asymmetric visual weights based on motion-risk mapping. Finally, a fusion-driven continuous state modulation solver, utilizing Bézier and Sigmoid functions, is introduced to eliminate multi-source asynchronous jitter and ensure C1 continuity. Experimental results demonstrate that the proposed multimodal fusion framework, by exploiting the known 80–120 ms physiological precedence of eye movements over head rotation, enables approximately 100 ms of anticipatory visual modulation, significantly outperforming baseline single-source methods. The approach yields a 27% reduction in Simulator Sickness Questionnaire (SSQ) scores, improves physiological comfort (an 18% increase in HRV), and drastically reduces mask perceptibility to 34%, all without sacrificing task performance. The findings provide an effective multimodal information fusion paradigm for human-in-the-loop immersive systems.

Keywords: multimodal sensor fusion, spatio-temporal information, intent prediction, dynamic visual modulation, virtual reality, human-in-the-loop systems.

Academic Editor:
Guoyan Wang

Submitted: 22 December 2025
Revised: 07 June 2026
Accepted: 23 June 2026
Published: 26 September 2026

Vol. 3, **No.** 3, 2026.
10.62762/CJIF.2025.472730

***Corresponding author:**
✉ Wenli Shang
rct.shanV2@bsu.by

1 Introduction

With the rapid development of virtual reality (VR) technologies, users can experience highly realistic

Citation
Shang, W., & Alena, K. (2026). Adaptive Field-of-View Restriction Based on Sigmoid and Bézier Smoothing Control Functions. *Chinese Journal of Information Fusion*, 3(3), 178–188.

visual and interactive environments. Nevertheless, visually induced motion sickness (VIMS) continues to pose a major challenge in human-in-the-loop immersive systems. According to the sensory conflict theory [1, 14], VIMS essentially arises from the spatio-temporal mismatch between multi-sensory inputs—specifically, when the motion cues provided by the visual system diverge from the self-motion signals detected by the vestibular system. Traditional mitigation techniques, such as Dynamic Field-of-View (DFOV) restriction [2], automatically narrow the visual field during user movement to suppress peripheral optic flow. Subsequently, content-based adaptive control methods have attempted to adjust visual parameters based on optic-flow density or head-motion states. However, these approaches primarily rely on single-modality kinematic sensors (e.g., IMU angular velocity), which inherently suffer from mechanical response delays and fail to capture the complex physiological and environmental contexts. Consequently, three critical limitations remain in existing visual modulation frameworks:

1) Lack of Cross-Modal Predictive Fusion: Conventional methods rely solely on post-action angular-velocity thresholds. By ignoring the physiological anticipatory signals (such as gaze intent), this single-sensor reliance leads to significant delayed activation that mismatches user intent.

2) Inadequate Spatial Data Integration: Excessive and global FOV shrinkage ignores the spatial distribution of visual risks, concurrently weakening presence and environmental consistency.

3) Temporal Discontinuity in State Modulation: Current algorithms predominantly adopt linear or piecewise-linear contraction. When processing dynamic sensor data, linear mapping often produces step-like control outputs, causing abrupt luminance changes and perceptual flicker at boundaries.

To address the bottlenecks of single-sensor reliance and discrete control, this paper proposes a novel Spatio-Temporal Information Fusion Framework for adaptive visual modulation in VR. The proposed framework deeply integrates multi-source heterogeneous data: high-frequency eye-tracking signals (representing physiological intent), IMU data (representing physical kinematics), and dense optic flow (representing environmental spatial features). This multi-source design is grounded in evidence that the human perceptual system integrates heterogeneous sensory cues in a statistically optimal, reliability-weighted manner [12], which motivates weighting each modality by its momentary informativeness rather than relying on any single sensor. Furthermore, instead of using discrete linear mapping, we introduce a fusion-driven continuous state modulation solver utilizing Sigmoid and Bézier functions. This mathematical solver ensures continuous state transitions (C1 continuity) to process the fused multidimensional data, effectively eliminating fusion-induced visual artifacts.

The primary contributions of this work, from an information fusion perspective, are as follows:

1)A cross-modal temporal intent prediction fusion model that integrates gaze and head-motion data to initiate system response approximately 100 ms in advance;

2)A visual-kinesthetic spatial data fusion strategy that constructs an asymmetric risk map based on real-time dense optic-flow perception;

3)A fusion-driven continuous state solver based on Sigmoid and Bézier functions that eliminates state-transition artifacts associated with multi-source data mapping;

4)A systematic evaluation validating the fusion framework's ability to optimize the balance between motion sickness, physiological comfort, and system immersion.

2 Related Work

2.1 Dynamic Field-of-View Restrictors

Fernandes and Feiner [2] proposed a dynamic FOV restriction (FOVR) model based on velocity-triggered mechanisms, demonstrating that reducing peripheral vision effectively decreases motion-sickness intensity. Kim et al. [18], as well as Lim et al. [7] further introduced content-aware dynamic FOV control methods that estimate motion-sickness levels using optic-flow features and IMU signals, and dynamically adjust mask opacity to alleviate discomfort while preserving immersion [18].

Wu and Rosenberg [3] proposed an optic-flow–based adaptive FOVR method that selectively occludes regions with high optic-flow intensity by computing dense optical flow in real time, thereby enabling asymmetric dynamic masking. Their results showed significant reductions in motion-sickness symptoms across various scenarios, accompanied by improvements in perceived presence. Complementary empirical evidence was provided by Zielasko et al. [16], who systematically examined the relationship between dynamic FOV reduction parameters and subjective sickness measures in an HMD-based task, confirming that the degree and rate of FOV restriction are key factors governing user discomfort.

2.2 Gaze-Driven Predictive Modulation

Adhanom et al. [4] introduced a Foveated FOV Restrictor, in which the mask center dynamically follows the user's gaze, preserving visual clarity in the foveal region while attenuating peripheral stimulation. Implemented on the HTC Vive Pro Eye with real-time eye tracking, their study demonstrated that gaze-driven restriction significantly enhances visual naturalness and user comfort compared with static masking.

Patney et al. [5] proposed dynamically foveated rendering, which maintains high resolution in the central vision while progressively down sampling peripheral regions, thereby reducing both rendering cost and peripheral motion stimulation; however, this approach modifies rendering resolution rather than applying an explicit opacity mask, representing a complementary rather than equivalent strategy to FOV restriction. Carnegie and Rhee [6] employed a dynamic depth-of-field technique that automatically adjusts focal depth during rapid rotations, effectively reducing visual fatigue and perceptual mismatch, though this operates on focus cues rather than field-of-view geometry and thus addresses a distinct aspect of visual comfort.

2.3 Temporal Smoothing in Visual Interventions

Optic-flow magnitude has been widely recognized as a key predictor of motion-sickness severity. Kim et al. [18] proposed an optic-flow–gradient–based adaptive FOV method that prioritizes FOV contraction in regions with intense motion. Lim et al. [7] employed a logistic function to smoothly modulate mask luminance, achieving gradual transitions and improved visual consistency. Collectively, these studies indicate that controlling peripheral optic-flow intensity and ensuring smooth temporal transitions are critical mechanisms for motion-sickness mitigation.

Previous research has shown that eye movements typically precede head motion by approximately 80–120 ms. Li et al. [9] developed a rapid eye-movement calibration model based on smooth pursuit, reducing gaze-prediction error to 0.4°. Park et al. [19] reported that motion-sickness ratings are significantly reduced when camera rotation aligns with gaze direction. Lou et al. [11] further validated the importance of gaze signals for action prediction using a multimodal gaze-prediction model. Similarly, Chang et al. [10] demonstrated that specific gaze behaviors are strong predictors of cybersickness severity. These findings provide empirical support for the gaze-based predictive triggering mechanism proposed in this work.

According to sensory conflict theory [1, 14] discrepancies between visual and vestibular inputs constitute the primary cause of motion sickness. Sensory rearrangement theory suggests that gradually varying visual stimuli can reduce peak neural conflicts and alleviate discomfort [13]. A systematic review by Ang and Quarles [20] emphasized that the smoothness and predictability of FOV transitions are key determinants of user comfort. Consequently, controlling the rate and curvature of FOV changes using Sigmoid or Bézier functions is theoretically well grounded.

Recent advancements in 2023 and 2024 further emphasize real-time prediction methods, utilizing time-series eye tracking to predict cybersickness dynamically [21] and exploring user-activated dynamic FOV restriction techniques [22].

In summary, while existing studies have made significant progress, most visual modulation frameworks still rely on single-modality sensors, which inherently lead to delayed system responses. Furthermore, the reliance on discrete or linear state transitions often causes temporal discontinuity and visual flicker. Therefore, there is a critical need for a spatio-temporal fusion framework that effectively integrates cross-modal anticipation with continuous mathematical solvers to directly address these research gaps.

3 Methodology

This section presents a unified spatio-temporal information fusion framework for the proposed adaptive VR visual modulation system. The approach is modeled as a predictive fusion architecture consisting of three stages:

(1) a cross-modal temporal intent fusion module that anticipates user turning intent prior to motion onset;

(2) a fusion-driven continuous state modulation solver utilizing continuously differentiable functions to regulate the fused discrete outputs;

(3) a visual-kinesthetic spatial data fusion module that constructs peripheral motion-risk maps for asymmetric dynamic masking.

3.1 Spatio-Temporal Fusion Architecture

The system architecture, illustrated in Figure 1, operates as a discrete-time dynamic fusion system taking heterogeneous multimodal sensor inputs: IMU angular velocity ω_{imu} and eye-tracking data g_{eye} at time t_k. The data fusion pipeline synthesizes the final mask opacity map $M(x, y, t)$ via Unity post-processing. The multi-sensor fusion system operates at a high-frequency $f \geq 90$ Hz with a total latency $\tau < 20\,ms$. The pipeline comprises five functional modules:

1. Sensor Acquisition: Collects ω_{imu} and g_{eye} at time t_k.
2. Predictive Triggering (C2): Detects saccadic onset to initiate early contraction (t_{start}).
3. Optic-flow Weighting (C3): Computes spatial risk field $R(x, y)$ based on motion intensity.
4. Smoothing Control (C1): Generates the temporal scalar $S(t)$ via Bézier functions.
5. Rendering: Synthesizes the final mask opacity map $M(x, y, t)$ via Unity post-processing.

3.2 C1: Fusion-Driven Continuous State Modulation

When fusing multi-rate and multi-modal sensor data (e.g., triggering signals from gaze and continuous velocity from IMU), direct linear mapping often produces step-like discrete outputs, causing abrupt visual flickering. To address this, we formulate the fused state modulation as a trajectory optimization problem.

Let F(t) denote the continuous state output at time t during a transition period T. We define a Visual Discomfort Cost Function J to penalize abrupt changes in visual flux caused by asynchronous data fusion. The optimization objective minimizes the second derivative of the state:

$$J(F) = \int_0^T \left(\frac{d^2F(t)}{dt^2} \right)^2 dt \tag{1}$$

Baseline Linear Mapping: The standard linear interpolation during sensor fusion is defined as $F_{lin}(t) = F_{start} + \frac{t}{T}(F_{end} - F_{start})$. Its first derivative (velocity) is discontinuous at boundaries $t = 0$ and $t = T$. Consequently, the second derivative (acceleration) contains Dirac delta impulses:

$$\frac{d^2F_{lin}}{dt^2} = \delta(t) - \delta(t - T) \tag{2}$$

This theoretically infinite acceleration at boundaries produces the "start-stop" system shock.

Proposed Bézier State Solver: To ensure C^1 continuity (continuous velocity) and minimize J during data fusion, we propose a mathematical state solver based on the cubic Bézier curve. With normalized time $u = \frac{t}{T} \in [0, 1]$, the state solver $F_{bez}(t)$ is defined as:

$$F_{bez}(t) = F_{start} + (F_{end} - F_{start})B(u) \tag{3}$$

where $B(u)$ is defined by the Bernstein polynomials with control points P_0, P_1, P_2, P_3 to enforce zero initial and final velocity ($F'(0) = F'(T) = 0$):

$$B(u) = (1-u)^3P_0 + 3(1-u)^2uP_1 + 3(1-u)u^2P_2 + u^3P_3 \tag{4}$$

Theorem 1 (Smoothness and Boundedness). *The proposed Bézier state solver $F_{bez}(t)$ is C^1 continuous on $t \in (0, T)$ and its acceleration is strictly bounded (as visually compared in Fig. 2).*

Proof. The time derivative of the normalized Bézier curve with the chosen control points is

$$\frac{dF_{bez}}{dt} = \frac{6u(1-u)}{T}\Delta F \tag{5}$$

Unlike the linear model, the acceleration of the Bézier model is finite and bounded by $\pm 6\frac{\Delta F}{T^2}$ throughout the interval. This mathematical property minimizes the cost J, ensuring visually comfortable state transitions without luminance transients during multi-sensor data synchronization. □

3.3 C2: Cross-Modal Temporal Intent Fusion

To compensate for system latency inherent in single-sensor systems, we introduce a temporal fusion mechanism based on the physiological precedence of eye movement over head movement. We define the triggering state $T_{state} \in \{0, 1\}$ using a dual-threshold fusion logic. A combined event is triggered ($T_{state} = 1$) if and only if both the gaze velocity $\omega_{eye}(t)$ and the cumulative gaze saccade amplitude $\Delta g(t)$ exceed their respective thresholds τ_v and τ_d:

$$T_{state}(t) = \begin{cases} 1, & \text{if } |\omega_{eye}(t)| > \tau_v \text{ AND } \|\Delta g(t)\| > \tau_d \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

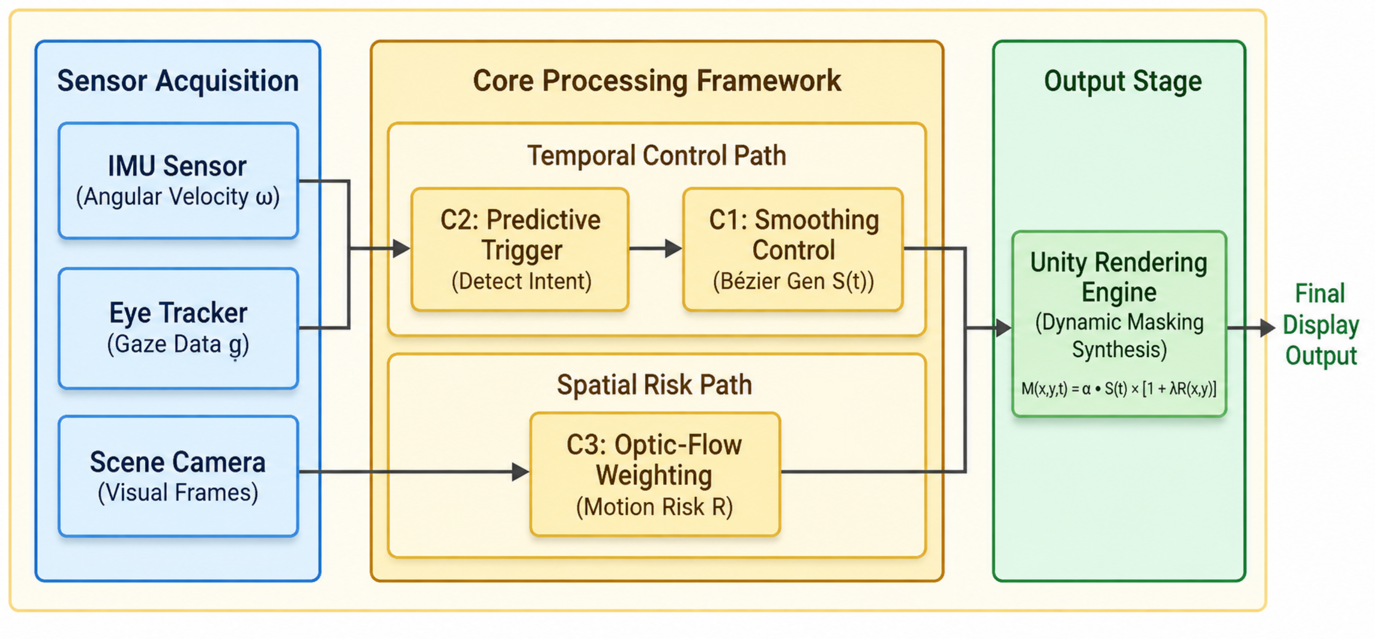

Figure 1. System Architecture.

We set the velocity threshold τ_v to $250°/s$ to reliably distinguish saccadic eye movements from smooth pursuit. According to the 'Main Sequence' relationship established by Bahill et al. [17], peak saccadic velocities typically exceed $200°/s$. This cross-modal fusion mechanism anticipates head rotation approximately 100 ms in advance.

3.4 C3: Visual-Kinesthetic Spatial Data Fusion

To construct the asymmetric modulation, we mathematically formulate the spatial risk distribution by fusing dense optic flow with motion intent. For a discretized screen grid $G = \{g_{ij}\}$, we define the Normalized Motion Risk Map R_{ij} based on the average flow magnitude $\bar{v}_{ij}$:

$$R_{ij} = \frac{\bar{v}_{ij} - v_{min}}{v_{max} - v_{min} + \epsilon} \quad (7)$$

The final fused visual state $M(x, y, t)$ is dynamically modulated by coupling the temporal state scalar $S(t)$ (derived from the Bézier solver) with the spatial risk weight $R(x, y)$. The data fusion blending equation is defined as:

$$M(x, y, t) = \alpha_{base} + S(t)[1 + \lambda R(x, y)] \quad (8)$$

This multi-dimensional fusion formulation ensures that the masking intervention is spatially optimized to suppress high-frequency motion stimuli, which are the primary contributors to motion sickness [8].

In our implementation, the core system parameters for the dynamic masking process were configured as follows: the unconstrained baseline field of view (F_{max}) was set to $110°$, and the maximum restriction limit (F_{min}) was capped at $70°$ to maintain necessary spatial awareness. The state recovery time, defined as the duration required to return from (F_{min}) to (F_{max}) after head motion ceases, was configured to 300 ms to ensure visual stability.

4 Experimental Design and Results

4.1 Experimental Objectives

The objective of this experiment is to compare the effects of three State modulation solvers during sensor fusion—Linear, Sigmoid, and Bézier—on motion-sickness discomfort, visual naturalness, and system performance. The study aims to verify whether the proposed smoothing control model can significantly alleviate motion sickness without compromising immersion or task performance.

4.2 Apparatus and Implementation

The experimental system was developed using the Unity 2021.3 LTS engine and deployed on an HTC Vive Pro Eye head-mounted display (HMD). The hardware setup and data flow are illustrated in Figure 3a, while the simulated virtual environment is shown in Figure 3b. The HMD features dual OLED displays with a combined resolution of 2880×1600 pixels (1440×1600 per eye), a refresh rate of 90 Hz, and a field of view (FOV) of $110°$. Integrated Tobii eye-tracking

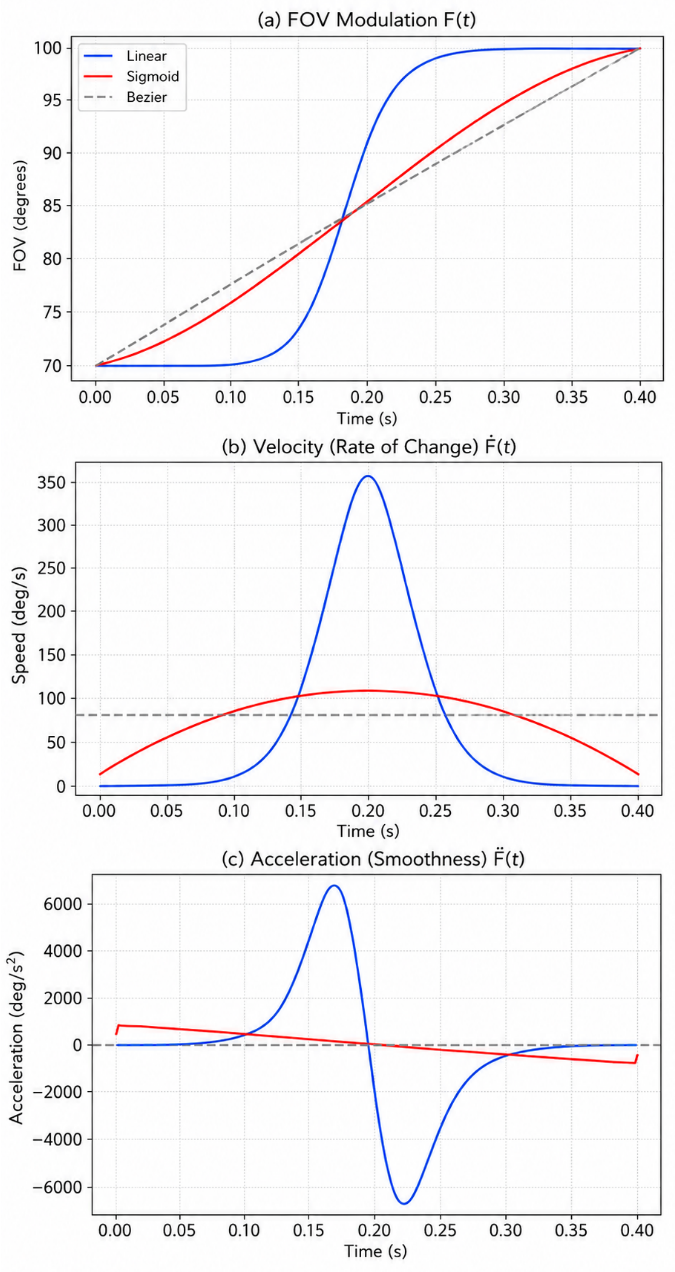

Figure 2. Kinematic comparison of different state solvers during multi-sensor data fusion. (a) Temporal profiles of visual state modulation $F(t)$ over a 400 ms transition. (b) Velocity profiles $\dot{F}(t)$. Note that the Baseline Linear method (gray dashed) requires instantaneous velocity jumps at start/end, which typically occurs when directly mapping asynchronous trigger signals. (c) Acceleration profiles $\ddot{F}(t)$ representing system jitter. The proposed Bézier solver (red) demonstrates continuous acceleration changes, thereby minimizing fusion-induced state shocks and avoiding the infinite impulse responses associated with linear multi-source mapping boundaries.

technology provided gaze data at a frequency of 120 Hz with an accuracy of $0.5^{\circ} - 1.1^{\circ}$, ensuring precise input for the predictive triggering mechanism. The VR simulation was executed on a laptop computer equipped with an Intel Core Ultra 9 285H processor and 32 GB RAM. Heart rate variability (HRV) data was collected using a Polar H10 sensor via Bluetooth.

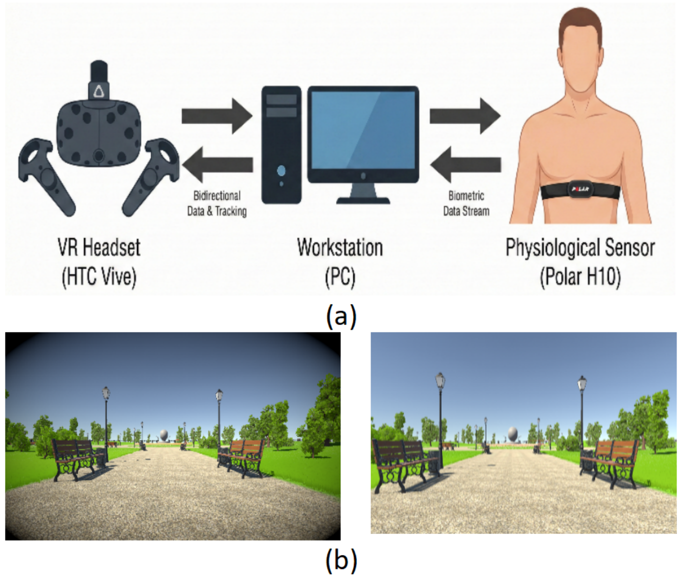

Figure 3. System implementation. (a) Schematic of the hardware apparatus and data flow between the VR headset, workstation, and physiological sensors. (b) The experimental environment in Unity, illustrating the visual intervention.

The top panel shows the unrestricted field of view (FOV), while the bottom panel demonstrates the adaptive masking triggered during head rotation to mitigate motion sickness.

4.3 Experimental Procedure

The experiment followed a within-subject design. Participants completed two distinct tasks under three control conditions (Linear, Sigmoid, Bézier) presented in a counterbalanced order to minimize learning effects. The workflow and logic for both tasks are illustrated in Figure 4. Each session began with a 5-point eye-tracking calibration followed by a 2-minute acclimatization period.

Task 1: Target Search Task This task assessed spatial perception and the impact of FOV restriction on visual scanning.

1. **Initialization:** The participant stood in the center of a virtual room.
2. **Stimulus:** A red luminous target (radius: 0.5m) appeared at random spherical coordinates (θ, ϕ) within the 360° panoramic environment.
3. **Action:** Participants were instructed to locate and fixate their gaze on the target as quickly

Table 1. Quantitative Comparison of Subjective and Objective Metrics across Different Solver Conditions.

Fusion State Solver Strategy	SSQ Reduction	Visual Naturalness (1–7)	HRV Increase (%)	Mask Perceptibility (%)	Frame Rate (Hz)
Linear	—	4.5	—	90	88.7
Sigmoid	↓ 22%	5.8	+15%	52	89.5
Bézier	↓ 27%	6.2	+18%	34	90.0

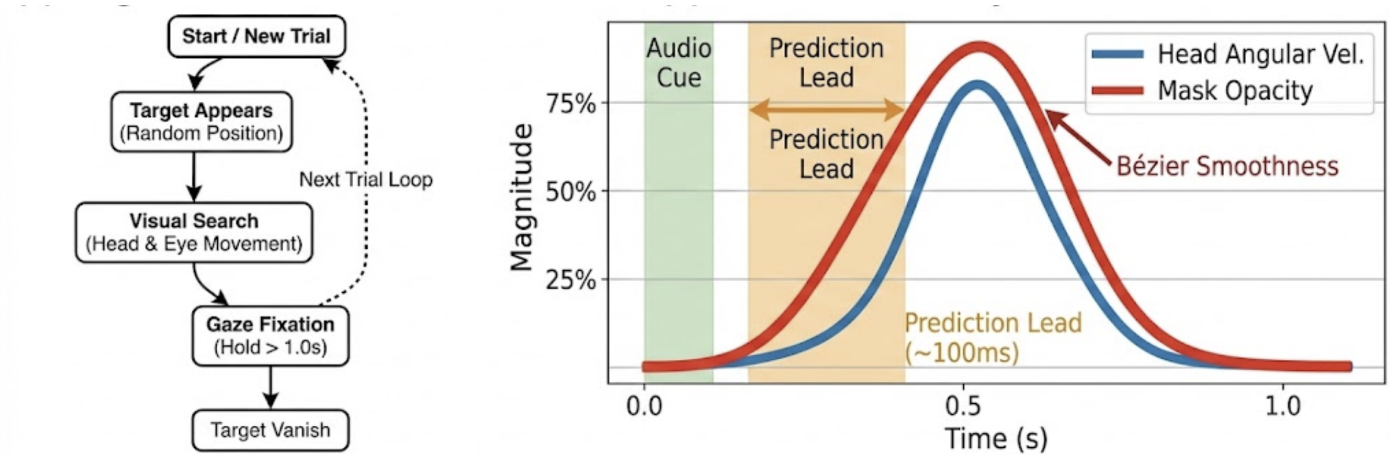

Figure 4. Schematic of the experimental tasks. (Left) Target Search Task logic flow. (Right) Head-Turn Tracking Task timeline showing the alignment of cues and mask activation.

as possible. Upon fixation for 1.0s, the target vanished, and a new one appeared.

4. **Duration:** The task lasted for 3 minutes per condition.

Task 2: Head-Turn Tracking Task This task evaluated motion comfort during rapid rotational movements, which are known to induce VIMS.

1. **Cue:** A visual arrow and 3D spatial audio cue prompted the participant to rotate their head left or right by 90°.
2. **Rotation:** The participant performed a rapid head rotation (peak velocity $> 150°/s$) to align with the designated direction.
3. **Rest:** A 2-second dwell time was required before the next cue.
4. **Duration:** The task involved 40 trials per condition (approx. 4 minutes).

4.4 Participants

Thirty healthy volunteers (15 male, 15 female; mean age: 23.4 ± 2.1 years) were recruited from the university. All participants had normal or corrected-to-normal vision and no history of vestibular disorders. Written informed consent was obtained prior to the experiment.

4.5 Measurements

Subjective Metrics:
Simulator Sickness Questionnaire (SSQ): Administered before and immediately after each condition to calculate the Total Severity Score (TSS).
Visual Naturalness: Participants rated the smoothness of the visual experience on a 7-point Likert scale (1 = extremely artificial/flickering, 7 = completely natural).
Mask Perceptibility: Participants estimated the percentage of time they were consciously aware of the black edges of the FOV mask during the task.

Physiological Metrics:
Heart Rate Variability (HRV): The standard deviation of NN intervals (SDNN) was calculated to quantify autonomic nervous system stress (higher SDNN indicates lower stress).

System Metrics:
Frame Rate (FPS): Logged to ensure the rendering performance remained above 90 Hz.

4.6 Results and Analysis

Subjective Results (SSQ & Naturalness): As presented in Table 1 and Figure 5, the proposed control methods significantly outperformed the Linear baseline. The Bézier method achieved the highest reduction in motion sickness, lowering the SSQ score by 27%,

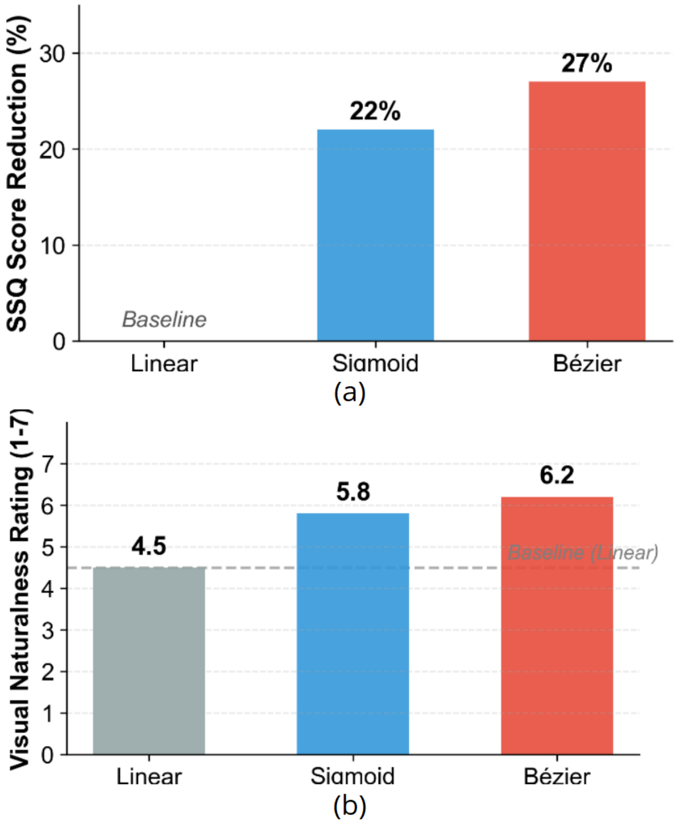

Figure 5. Comparison of subjective evaluation results across control conditions. (a) Reduction in Simulator Sickness Questionnaire (SSQ) scores relative to the Linear baseline. (b) Visual Naturalness ratings on a 7-point Likert scale.

compared to a 22% reduction with the Sigmoid method.

Crucially, the Bézier method demonstrated superior visual quality. Participants rated its Visual Naturalness significantly higher (6.2/7.0) than the Linear baseline (4.5/7.0). Furthermore, the Mask Perceptibility—the percentage of time users consciously noticed the black edges—dropped drastically from 90% (Linear) to 34% (Bézier). This indicates that the smooth acceleration profile of the Bézier curve effectively makes the FOV restriction "invisible" to the user, thereby preserving immersion.

Physiological & System Results: Physiological metrics supported the subjective findings. HRV (SDNN) increased by 18% under the Bézier condition, reflecting a measurable decrease in physiological stress compared to the baseline.

In terms of system performance, the average Frame Rate remained stable at approx. 90 Hz (Linear: 88.7 Hz, Sigmoid: 89.5 Hz, Bézier: 90.0 Hz) across all conditions. This confirms that the computational cost of the proposed smoothing algorithm is negligible and fully satisfies the real-time rendering requirements of VR applications.

5 Discussion

5.1 Mechanisms of Continuous State Solvers in Fusion

When fusing asynchronous, multi-rate sensor data (e.g., discrete gaze-trigger events with continuous IMU streams), linear mapping inevitably introduces temporal discontinuities in the system output. According to sensory rearrangement theory, the human nervous system exhibits greater adaptability to continuous stimuli than to abrupt state changes. By regulating the first- and second-order derivatives of the fused visual states, the Sigmoid and Bézier solvers effectively eliminate the multi-source fusion jitter, matching physiological visual response profiles more naturally than traditional discrete controllers.

5.2 Temporal Advantages of Cross-Modal Prediction

Traditional single-sensor models, such as the IMU-velocity-triggered model proposed by Fernandes and Feiner [2], inherently suffer from mechanical and communication delays of 50–100 ms. In contrast, the cross-modal temporal fusion mechanism employed in this study integrates physiological signals (eye-tracking) with physical kinematics (IMU) to enable anticipatory triggering. This fusion strategy ensures that the visual modulation completes prior to the head rotation, successfully synchronizing the visual rendering pipeline with the user's vestibular cues.

5.3 Visual-Kinesthetic Spatial Data Fusion

Building upon the insight that regions with high optic-flow intensity correspond to elevated motion-sickness risk , our proposed spatial fusion factor selectively maps the kinesthetic intent onto the environmental visual features. Results demonstrate that this spatio-temporal fusion strategy asymmetricly enhances masking in high-risk peripheral regions while preserving central visual clarity and task accuracy.

5.4 Contribution of Individual Modules

While the current experimental design evaluates the integrated framework, theoretical analysis and subjective metrics suggest distinct contributions from each module. The Bézier smoothing solver contributes primarily to the improvement in Visual

Naturalness (rated 6.2/7.0) and to the reduction in Mask Perceptibility (which dropped to 34%). Conversely, the cross-modal gaze prediction and optic-flow spatial weighting are the primary drivers for the 27% reduction in SSQ scores, as they directly address the temporal delay and spatial localization of motion-sickness triggers. Future studies will incorporate comprehensive ablation experiments to quantify the isolated contribution of each module.

5.5 Comparison with Related Fusion Frameworks

Compared with foveated restrictors driven by single gaze inputs , the present multimodal framework achieves comparable visual naturalness while offering superior temporal responsiveness through intent fusion. Relative to logistic luminance modulation approaches , the proposed method leverages mathematically continuous solvers to handle multidimensional data, resulting in more stable transitions and reducing the system shocks typically associated with multi-sensor integration.

5.6 Theoretical Implications for Human-in-the-Loop Systems

From an information fusion perspective, this work establishes a unified framework integrating temporal anticipation (cross-modal intent fusion) with spatial adaptability (optic-flow mapping) and continuous state rendering. Despite these promising results, several avenues for future optimization remain:

- **Reinforcement learning-based fusion weight adaptation:** Deep reinforcement learning models could be employed to automatically adjust the fusion weights and Bézier control points by minimizing predicted motion-sickness scores derived from physiological data streams (e.g., HRV, Electrodermal Activity [EDA]).
- **Multi-agent collaborative spatial fusion:** In shared immersive environments, multi-user gaze and trajectory data could be fused to predict group attention and coordinate global visual optimization.
- **Edge-assisted lightweight fusion:** Combining this spatio-temporal framework with edge-assisted computing strategies may enable low-latency, multi-sensor deployment on standalone wireless headsets. For instance, deep reinforcement learning-based rendering offloading and subband allocation techniques for edge-assisted VR systems [15] could be integrated with the proposed fusion pipeline to further reduce end-to-end latency under constrained wireless bandwidth.

6 Conclusion

This paper presents a novel spatio-temporal information fusion framework designed to overcome the limitations of single-sensor reliance in VR visual modulation. By deeply integrating cross-modal intent prediction (gaze and IMU), spatial feature fusion (dense optic flow), and a continuous state mathematical solver (Bézier/Sigmoid), the proposed system achieves natural, anticipatory, and jitter-free multi-source data synchronization. Experimental results confirm that this multimodal fusion approach significantly outperforms conventional single-source linear models, yielding a 27% reduction in motion sickness and preserving high-fidelity immersion (reducing mask perceptibility to 34%). The study validates the critical role of spatio-temporal data fusion and continuous state solving in optimizing human-in-the-loop immersive systems.

Despite the promising results, this study has certain limitations. First, the fixed fusion weights may not perfectly accommodate individual differences in motion sickness susceptibility. Second, the current evaluation was conducted in single-user scenarios, which limits the exploration of multi-agent visual optimization. Future work will focus on integrating deep reinforcement learning to dynamically adapt fusion parameters based on real-time physiological feedback and extending the framework to multi-user collaborative environments.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

The experiments in this study involved non-invasive commercial VR and physiological equipment (HTC Vive Pro Eye and Polar H10). According to the institutional guidelines of Belarusian State University, ethics review for this type of non-invasive, low-risk behavioral study was exempted. All participants were informed about the experimental procedure and potential transient effects (mild motion sickness), and written informed consent was obtained before participation.

References

[1] Oman, C. M. (1990). Motion sickness: a synthesis and evaluation of the sensory conflict theory. *Canadian journal of physiology and pharmacology*, *68*(2), 294-303. [CrossRef]

[2] Fernandes, A. S., & Feiner, S. K. (2016, March). Combating VR sickness through subtle dynamic field-of-view modification. In *2016 IEEE symposium on 3D user interfaces (3DUI)* (pp. 201-210). IEEE. [CrossRef]

[3] Wu, F., & Suma Rosenberg, E. (2022, November). Adaptive field-of-view restriction: Limiting optical flow to mitigate cybersickness in virtual reality. In *Proceedings of the 28th ACM symposium on virtual reality software and technology* (pp. 1-11). [CrossRef]

[4] Adhanom, I. B., Griffin, N. N., MacNeilage, P., & Folmer, E. (2020, March). The effect of a foveated field-of-view restrictor on VR sickness. In *2020 IEEE conference on virtual reality and 3D user interfaces (VR)* (pp. 645-652). IEEE. [CrossRef]

[5] Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., ... & Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions On Graphics (TOG)*, *35*(6), 1-12. [CrossRef]

[6] Carnegie, K., & Rhee, T. (2015). Reducing visual discomfort with HMDs using dynamic depth of field. *IEEE computer graphics and applications*, *35*(5), 34-41. [CrossRef]

[7] Lim, K., Lee, J., Won, K., Kala, N., & Lee, T. (2021). A novel method for VR sickness reduction based on dynamic field of view processing. *Virtual Reality*, *25*(2), 331-340. [CrossRef]

[8] So, R. H., Lo, W. T., & Ho, A. T. (2001). Effects of navigation speed on motion sickness caused by an immersive virtual environment. *Human factors*, *43*(3), 452-461. [CrossRef]

[9] Li, Y., Guo, L., Sun, G., Fu, R., Yan, Z., & Liang, J. (2022). Eye tracking calibration based on smooth pursuit with regulated visual guidance. In *Proceedings of the 14th International Joint Conference on Computational Intelligence (IJCCI)* (pp. 417-425). [CrossRef]

[10] Chang, E., Kim, H. T., & Yoo, B. (2021). Predicting cybersickness based on user's gaze behaviors in HMD-based virtual reality. *Journal of Computational Design and Engineering*, *8*(2), 728-739. [CrossRef]

[11] Lou, Z., Cui, Q., Wang, H., Tang, X., & Zhou, H. (2024, June). Multimodal sense-informed forecasting of 3d human motions. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2144-2154). IEEE. [CrossRef]

[12] Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433. [CrossRef]

[13] Reason, J. T. (1978). Motion sickness adaptation: a neural mismatch model. *Journal of the royal society of medicine*, *71*(11), 819-829. [CrossRef]

[14] Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A., Rizzo, A., & Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis. *Frontiers in human neuroscience*, *14*, 96. [CrossRef]

[15] Iftikar, M., Asiedu, D. K. P., Nishio, T., & Yun, J. H. (2024). Deep reinforcement learning-based overfill rendering, offloading, and subband allocation for edge-assisted VR system. *IEEE Access*, *12*, 149147-149161. [CrossRef]

[16] Zielasko, D., Meißner, A., Freitag, S., Weyers, B., & Kuhlen, T. W. (2024). Dynamic field of view reduction related to subjective sickness measures in an HMD-based data analysis task. *arXiv preprint arXiv:2403.07992*. [CrossRef]

[17] Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical biosciences*, *24*(3-4), 191-204. [CrossRef]

[18] Kim, S., Lee, S., Kala, N., Lee, J., & Choe, W. (2018). An effective FoV restriction approach to mitigate VR sickness on mobile devices. *Journal of the Society for Information Display*, *26*(6), 376-384. [CrossRef]

[19] Park, M. H., Yun, K., & Kim, G. J. (2023). Reducing VR sickness by directing user gaze to motion singularity point/region as effective rest frame. *IEEE Access*, *11*, 34227-34237. [CrossRef]

[20] Ang, S., & Quarles, J. (2023). Reduction of cybersickness in head mounted displays use: A systematic review and taxonomy of current strategies. *Frontiers in Virtual reality*, *4*, 1027552. [CrossRef]

[21] Shimada, S., Ikei, Y., Nishiuchi, N., & Yem, V. (2023, March). Study of cybersickness prediction in real time using eye tracking data. In *2023 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)* (pp. 871-872). IEEE. [CrossRef]

[22] Yang, Y., Kim, H., & Kim, G. J. (2024, March). u-DFOV: User-activated dynamic field of view restriction for managing cybersickness and task performance. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 729-730). IEEE. [CrossRef]

Wenli Shang is a PhD candidate at the Faculty of Radiophysics and Computer Technologies, Belarusian State University. Her academic background is focused on computer technologies and applied research in immersive systems. Her current research interests include virtual reality–induced cybersickness and its mitigation strategies. She also works on the design and development of social agents in Unity-based virtual environments. Her work aims to improve user experience and interaction quality in immersive virtual reality systems. (Email: rct.shanV2@bsu.by)

Kazlova Alena is an Associate Professor and Head of Department at Belarusian State University. She holds a PhD in Physics and Mathematics and has extensive experience in academic research and teaching. Her research interests include information technologies and intelligent data processing methods. She also works on neural network–based and fuzzy modeling approaches for scientific applications. Her work focuses on the analysis and processing of data from physical and engineering systems. (Email: kozlova@bsu.by)