Data Engineering and Applications

Research Paper

Robust Visual SLAM Framework Based on Human Saccadic Eye Movement Modeling

Mustafa Demir[1,*] and Tülay Yaman[1]

[1] Faculty of Computer Engineering, Koc University, Istanbul, 34450, Turkey
*Corresponding author: mustafa.de@ku.edu.tr

Abstract. Robust visual simultaneous localization and mapping (SLAM) is fundamental for robotic navigation in unfamiliar or dynamic environments; however, traditional visual SLAM systems often suffer from accuracy and stability losses due to dynamic objects, occlusions, and abrupt scene changes. To address these challenges, we propose a novel SLAM framework inspired by human saccadic eye movements, which incorporates a neurobiologically-based inhibition-of-return mechanism to suppress redundant operations, adaptively allocate computational resources, and emphasize salient information. The proposed system embeds saccade-driven attention directly within the SLAM process, integrating adaptive map maintenance and attention-based real-time feature selection. Experiments conducted across diverse indoor, outdoor, and semi-structured environments—including complex scenes with dynamic traffic—demonstrate that our approach significantly improves performance. Specifically, the saccade-inspired model achieves an absolute trajectory error (ATE) of 0.139 m in dynamic outdoor sequences, outperforming baseline methods such as ORB-SLAM3 (0.176 m) and VINS (0.168 m), and maintains a median feature retention ratio of 69.2% over 3000-frame indoor sequences (vs. 54.6% and 48.1% for baselines). Robustness tests under occlusion and sensor disturbance confirm stable localization and faster recovery. These results validate the effectiveness of human-inspired attention mechanisms for enhancing SLAM robustness, precision, and resource efficiency in real-world, dynamic conditions.

Keywords: *Visual Perception, Simultaneous Localization and Mapping, Attention Mechanism, Robotics*

Received on 30 October 2024, Accepted on 08 May 2025, Published on 17 May 2025

Introduction

In modern robotic systems, such as autonomous driving, mobile robot patrol, production line inspection work, etc., visual simultaneous localisation and map-building (SLAM) has emerged as the fundamental technique for achieving motion estimates [1,2]. Recent developments in vision-based SLAM have reduced the requirement for additional facilities like cameras by offering more reliable and quick systems for portable devices. SLAM systems continue to perform poorly in situations like dynamic barriers, abrupt lighting changes, occlusion of the environment, and high levels of sensor noise. Their practical application under uncertainty is severely limited because these challenges frequently result in tracking failures, map corruptions, and loss of localization [3,4]. An intelligence- and adaptation-driven perceptual strategy must be used since traditional methods of static feature extraction or resource allocation are unsuitable for changing settings due to their lack of adaptability [5,6].

Humans are thought to use saccadic eye movement as an effective attention-sampling technique to quickly and reliably comprehend complex scenes across many scales, according to neuroscientific and psychophysical research [7]. In an adversarial visual environment, saccade behaviour decreases distraction by suppressing irrelevant inputs, allowing Biological Vision to swiftly shift its focus and enhance Cognitive Resource utilisation [8]. We provide a different SLAM system that incorporates a saccade-inspired selective attention and dynamic resource selection mechanism based on the evolutionary knowledge mentioned above. We describe how to create an Adaptive Feature Priority and Resource Budgeting module, formalise the saccadic Sampling process

under SLAM, and include them into a traditional visual SLAM system [9,10]. The suggested approach, which uses a mathematical model of the scene-salience distribution, can simultaneously carry out feature selection and computationally load division to enhance the accuracy and completeness of maps that adapt to complex scenes. Comprehensive studies on challenging bespoke datasets and public datasets have confirmed these claims [11].

The structure of this document is as follows. examining pertinent studies on bioinspired perception, attention modelling, and robust visual SLAM. This part will establish a mathematical-neurobiological foundation to translate the theory of saccadic motions into computations. The design and implementation of the Saccade-Driven Robust SLAM system are presented in Section 4. Comprehensive experimentally assessed contents and analyses are presented in Section 5: Ablation Studies. Section 6 wraps up and suggests future research directions.

Related Work

Visual SLAM Overview and Robustness Challenges

Some crucial developments, including as ORB-SLAM, DS-OS, and VINS, which all offer reliable baselines for monocular and multiple-sensor navigation and map-building in unfamiliar situations, have propelled the development of visual simultaneous localisation and mapping (SLAM). Due to these systems, visual SLAM is being used in mobile robots, drones, AR/VR technologies, and other real-world applications. These models operate exceptionally well in perfect or reasonably challenging conditions, but they struggle in dynamic, congested, and dimly lit environments [14,15]. Conventional features have a tendency to degrade quickly in situations like changing textures and motions, and they can't reliably identify when some aspects of the feature information vanish as a result of movement or shifting environmental influences. Semantic filtering, object-level tracking, and adaptive mapping are some methods to lower computation costs, however they will get more complicated and fail in dynamic environments. Furthermore, situations with unavoidable illumination changes, recurring patterns, and a disorganised setting will create uncertainty, lower trajectory estimation accuracy, and decrease map reliability.

Visual Attention, Saccadic Mechanisms, and Their Computational Models

The aforementioned research lines, which examine how people effectively resolve such robustness problems in vision, are typically motivated by the discipline of cognitive neuroscience. To put it another way, a brief blink of the eyes is intended to move the visual focus closer to the parts of the field of view that are most likely to draw attention. Saliency models and attention-guided feature selection in vision tasks are computational representations of the underlying biological phenomena. The robustness of the vision system to clutter and distractions can be enhanced by saliency-based frameworks, which prioritise the order of spatial or task-relevance [16,17]. Numerous computational attention models have been proposed, but those that combine top-down task cues with bottom-up stimulus saliency are especially good at replicating saccade-selection processes. These models are rarely used in conjunction with geometric estimates and SLAMB-ackends, despite the fact that they can improve recognition and track robustness in certain situations. The majority of existing approaches lack dynamic interactions with the mapping and positioning iteration mechanism and merely view attention as a preliminary processing step for keypoint detection or region proposal [18].

Bio-inspired and Neuromorphic Approaches in Perception and SLAM

A growing number of researchers have focused on bioinspired perception for robots in recent years, attempting to incorporate the efficient and adaptive principles found in nature into robot vision SLAM systems. The event-triggered and high dynamic range feature of the neuromorphic vision sensor, which mimics the biological retina, has drawn interest [19]. By using integration techniques, a number of investigators of neuromorphic data streams and SLAM have been able to improve time-domain accuracy and stability. Other research avenues for designing SLAM systems include neural networks, recurrent structures, and evolution algorithms. The fundamental mechanisms of saccadic attention, such as hierarchical prioritisation, context-sensitive feature gating, and active resource allocation, have not yet been fully embedded by these efforts, despite some progress in improving the adaptability and energy efficiency of SLAM system scenes. Few bio-inspired techniques are integrated as appendices but not included in basic geometric and probabilistic reasoning systems for SLAM, indicating the persistence of serious integration issues on both the algorithm and system sides [20]. These

shortcomings point to a crucial area for development: creating a generic model of saccadic attention to direct feature filtering, information processing path creation, and computation scheduling throughout the whole SLAM system, rather than just as a temporary assistance mechanism. [21] The subsequent work begins with a unified method.

Modeling Human Saccadic Eye Movements

Neurobiological Principles

Extremely efficient and adaptable reactions to various stimuli, rendering them appropriate for investigation by anyone examining artificial vision and perception. The core of the efficiency issue is a saccadic eye movement phenomenon: rapid and accurate motions of the eyeball to reposition the fovea onto a new target area within the visual field. The monkey visual system operates in stages, employing sequences of saccades and fixations instead of continuously generating a comprehensive global map of the environment. Saccades typically last between 20 to 200 milliseconds; they rapidly shift focus to various elements within the visual field or enhance clarity on objects requiring heightened attention [22].

Saccadic behaviour is orchestrated by the superior colliculus and fronto-eye regions of the brain, integrating bottom-up sensory inputs with top-down cognitive anticipations. Salient stimuli, including sudden movements and significant brightness contrasts, receive a transient fixation preference. Subsequently, by the weighted aggregation of temporal brain activity state data within constrained intervals, a restricted subset is selected with precision that adapts rapidly. During a saccade, visual perception is temporarily suppressed (saccadic masking), resulting in diminished attention to indistinct or ambiguous stimuli to alleviate cognitive strain.

Sparse Sampling's active learning technique demonstrates significant efficacy in non-incremental Robust Perception by swiftly eliminating extraneous data from view updates, promptly assessing feature weights through noise source filtration, and optimally achieving a balance among precision, recall, and memory expenditures. Approximately 3 to 5 saccades per second occur in reaction to contextual cues, variations in localised target features, and other environmental circumstances; nonetheless, confidence is elusive. This neurobiological protocol of selective and repetitive attention reveals a discernible deficiency in scene understanding that artificial systems lack. Consequently, these fundamental insights provide direction for creating computer models that emulate the resilience of saccadic-driven vision in artificial visual systems.

Mathematical Model

Translating saccadic eye movement into a computational model requires precise abstraction of both spatial and temporal mechanisms. In the visual system, gaze shifts are governed by the interplay of feature-driven saliency, contextual inhibition, and adaptive prioritization-a process that can be formalized mathematically for integration into artificial perception pipelines.

Let $S(t)$ denote the saliency map at time t, encoding the likelihood that each region within the visual field will attract the next fixation. The probability $P_{fix}(x,t)$ of selecting spatial position x as the saccadic target is given by

$$P_{fix}(x,t) = \frac{S(x,t) \cdot W(x,t)}{\int_{\Omega} S(y,t) \cdot W(y,t)\, dy} \quad \text{Eq. (1)}$$

where $W(x,t)$ is a dynamic weighting term, reducing the re-selection of previously attended or contextually less relevant zones, and Ω covers the current scene. This formulation inherently balances bottom-up saliency and inhibition-of-return, promoting efficient and diverse information sampling.

The temporal profile of a saccade can be described by its velocity profile. The instantaneous eye speed $v(t)$ during movement period $[t_0, t_1]$ is typically modeled as

https://doi.org/10.64972/dea.2025.v4i2.2439d:117-128

$$v(t) = v_{\max} \cdot \exp\left(-\frac{(t - t_{\text{mid}})^2}{2\sigma^2}\right) \quad \text{Eq. (2)}$$

where $v_{\max ?}$ is the saccade's peak velocity, $t_{\text{mid}?}$ locates the velocity peak, and σ reflects movement duration. Empirically, peak velocity grows with saccade amplitude following

$$v_{\max} = k \cdot A^{\alpha} \quad \text{Eq. (3)}$$

with A the saccade amplitude and k, α empirically set constants consistent with the main sequence relationship observed in biological vision.

Collectively, these elements enable modeling the sequence of fixations and saccades as a Markov process, where visual input, current fixation, and an evolving saliency landscape dynamically determine the next gaze position. This iterative mechanism, succinctly visualized in Figure 1, begins from raw scene input, flows through dynamic saliency and attention-based selection, and culminates in saccade generation and fixation update.

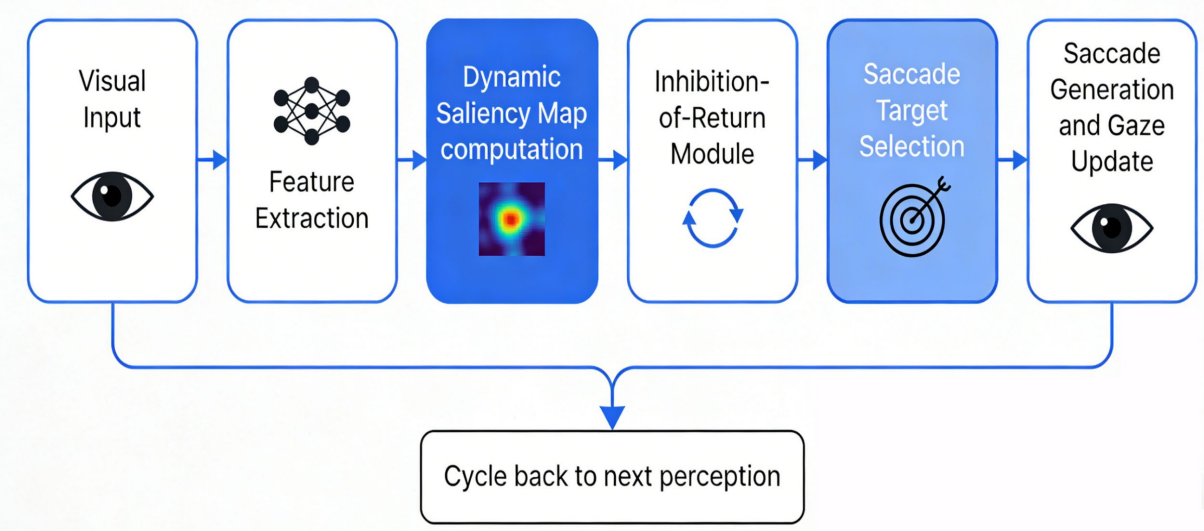

Figure 1. Saccadic Eye Movement Perception Model.

A schematic flowchart depicting the perception pipeline: (1) visual input and feature extraction; (2) dynamic saliency and inhibition-driven selection; and (3) saccade generation and gaze update. Embedding such mathematical rigor into artificial perceptual frameworks fundamentally equips engineered systems with the high-speed, context-sensitive attention and selectivity that underpin robust human vision.

Implications for Visual Perception

The design of a dependable Visual-SLAM system is directly impacted by the computation modelling of the saccadic mechanism. The SLAM front-end is guided by a saliency-driven fixation mechanism to more precisely concentrate its computation at areas with higher information rather than everywhere. In light of these issues, this work aims to reduce clutter, motion blur, and other issues in order to increase tracking accuracy and guarantee stability.

The weight function in the chosen formula is paired with the inhibitory return mechanism to prevent excessive repetition of visiting recently visited regions. They both extend their uses by reducing calculation to some degree; short-term or incorrect information poses no such risk.

Saccadic modelling also aligns with the SLAM pipeline's repetitive and serial structure. It adaptively chooses feature maps for localisation and mapping at each new frame based on the prior fixed state and updates the saliency distribution. For a longer duration of stable performance, feedback-driven updates improve adaptation in dynamic environments. To put it briefly, the SLAM system can be strengthened in dynamic and uncertain environment adaptation by including saccade-based attention models.

https://doi.org/10.64972/dea.2025.v4i2.2439d:117-128

Saccade-Inspired Robust Visual SLAM Framework

Overall System Architecture

In order to provide effective, precise visual odometry and map-building algorithms, the suggested saccade-driven SLAM architecture aims to mimic an adaptive-sampling and resource-allocation method of biologically inspired vision systems. Create an open architecture for this system that separates the core-SLAM calculation, feature-management, and perception-attention modules, respectively. There are certain relationships between the main components, as seen in Figure 2.

Continuously obtaining visual data through frames is the foundation for real-world perception. A real-time saliency map was created for every frame using both high-order contextual priors and low-level cues like intensity and motion. The attention mechanism is known as saccade selectors because it can offer a way to direct the choice of fixations. It minimises the number of computations by selecting regions for calculation based on projected information gain, in contrast to uniform sampling of features.

The Adaptive Feature Manager is directly connected to the Saccadic Selector Interface. The saliency region's features are merged with temporal attention weights, prioritised, and updated on a regular basis. These weights are sent over successive frames, and these systems are nevertheless able to maintain corresponding points in challenging situations with visual obstacles, light changes, etc. In order to expand Spatial Exploration and lessen Overfitting, Inhibitory-Return presents an improved version of Feature Pool by rejecting regions that have been studied more recently.

The SLAM front-end would then receive those chosen characteristics from the downstream portion of it for geometric position estimation and data-related processing. The front end is improved with a function that can modify its resources based on the current attention distribution; it allocates more processing power to high-priority areas, modifies temporal resolution or descriptor scale in accordance with requirements, etc. Failure detection mechanisms identify the loss of track confidence, etc., and quickly request a re-allocation of attention resources.

Pose Graph Optimisation, Loop Closure Detection, and Global Map Management are all part of the core SLAM Back-End. In order to facilitate context-aware localisation and improve long-term resilience, saliency and attention information are kept in an auxiliary way within the map representation. Real-time information updates for modifying the exploration-exploitation trade-off at high stress levels can be obtained through both channels of communication between the relevant attention/feature modules and the backend. An explicit systems-wide viewpoint is depicted as follows in Figure 2:

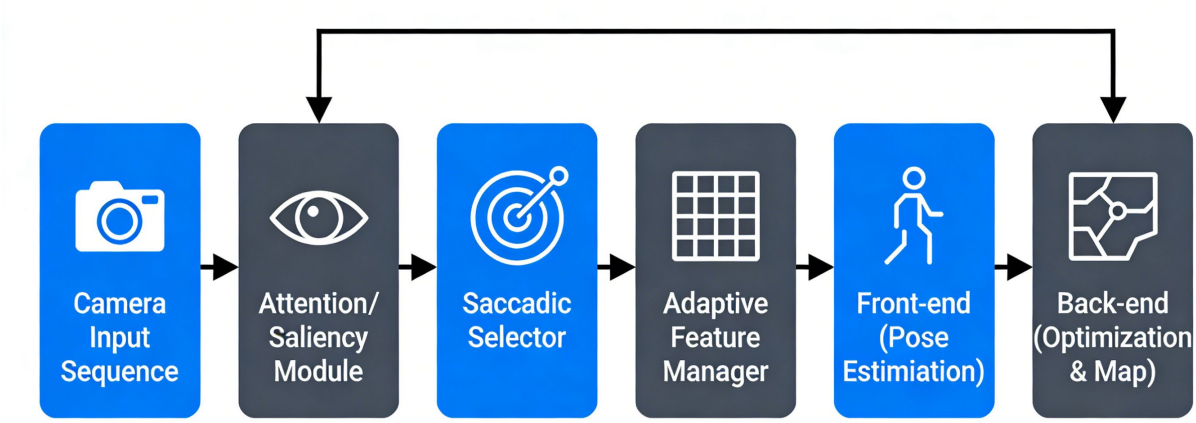

Figure 2. Saccade-Driven SLAM Framework Architecture.

This architecture fundamentally redefines the perception pipeline by embedding saccadic attention as a native principle of feature selection and computation steering, rather than as an external preprocessing step. The following sections detail the adaptive weighting, resource allocation, and the mechanisms for pipeline integration.

Adaptive Feature and Resource Strategy

Central to the saccade-inspired framework is the formulation of selective and adaptive mechanisms for feature prioritization and resource management. Traditional SLAM approaches distribute computational resources evenly across candidate features, which often leads to inefficiency and fragility in visually complex or dynamic environments. By explicitly modeling scene saliency and dynamically weighting feature importance, the proposed system is able to allocate attention and computation in a manner that closely mirrors the efficiency of biological perception.

Each visual frame is processed to generate a normalized saliency map, defined over pixel coordinates as $S(x,y,t)$, where t is the time index. Saliency at each location arises from a fusion of contrast, motion, and semantic priors, captured by the following formulation:

$$S(x,y,t)=\frac{C(x,y,t)+\beta_1 M(x,y,t)+\beta_2 P(x,y,t)}{Z} \quad \text{Eq. (4)}$$

Here, C quantifies local intensity or edge contrast, M encodes motion saliency, P introduces higher-level or task-centric priors, and β_1,β_2 are fusion coefficients. The term Z normalizes the response across the entire image, ensuring values are comparable and robust to scene variations.

Following saliency computation, all initially detected feature candidates $F_{\text{raw}?}=\{f_i\}$ are assigned selection probabilities according to their spatial saliency and additional factors such as temporal stability or inhibition of recent fixations. For each feature, the selection probability is formulated as:

$$p_i=\frac{S(x_i,y_i,t)\cdot w_i}{\sum_j S(x_j,y_j,t)\cdot w_j} \quad \text{Eq. (5)}$$

The variable w_i encodes both temporal reliability and the effect of inhibition-of-return, promoting the consistent use of stable landmarks while discouraging redundant sampling of recently attended locations. This soft probabilistic attention ensures that salient, distinctive, and spatially diverse features are prioritized for downstream SLAM processing.

In order to maximize computational efficiency, resource allocation-such as descriptor computation, matching, or optimization iterations-is adaptively distributed. For any image region R_k, the computational budget $\mathcal{R}_k$ is determined by the relative aggregate priority of its contained features:

$$\mathcal{R}_k=\mathcal{R}_{\max}\cdot\frac{\sum_{(x_i,y_i)\in R_k} p_i}{\sum_l \sum_{(x_j,y_j)\in R_l} p_j} \quad \text{Eq. (6)}$$

This enables processing load to intensify locally in areas with the highest expected informativeness or ambiguity, while reducing effort in predictable, low-value segmentsreplicating biological vision's apparent parsimony without compromising accuracy.

A key aspect that enables continual exploration is the inhibition-of-return (loR) mask, $\mathrm{IOR}(x,y,t)$, which assigns lower probabilities to locations previously fixated in the recent history. The final selection probability for each candidate becomes:

$$\grave{p}_i=p_i\cdot\left[1-\mathrm{IOR}(x_i,y_i,t)\right] \quad \text{Eq. (7)}$$

After each saccade and processing cycle, the loR mask is locally reinforced at selected locations and decays elsewhere, thereby balancing persistent exploration and stability over time.

The entire adaptive strategy is embedded natively in the SLAM perception pipeline, with realtime updates to all weights and budget allocations as the scene evolves. Data flows and information dependencies among saliency analysis, feature selection, and computational scheduling are visually depicted in Figure 2 in the previous section.

This fusion of selective attention, inhibition-driven novelty, and dynamic resource prioritization empowers the SLAM system to consistently focus on the most valuable scene content, enhancing both tracking robustness and operational efficiency even in complex or rapidly changing environments.

SLAM Pipeline Integration

For a seamless integration effect and excellent computation speed, the saccade-influenced attention model can readily join at any point in the visual-SLAM workflow. During front-end processing, the adaptive attention mechanism only permits pose estimation and feature tracking of items that are considered contextually significant and salient. In order to improve tracking accuracy by lowering failures brought on by dynamic scenes and cluttering, the adaptive probability weight function is employed to filter out raw features that are unstable, redundant, or uninformative during geometric estimation.

In the matching process, data association is achieved using an attention-guided weight update. In order to greatly reduce false positives and improve resilience against visual aliasing or discontinuous changes, features that are left over after the selection of those with both continuous saliency and inhibited-filtered scores are employed for inter-frame comparison. outlier-suppression using a technique that makes advantage of temporal attention profiles to boost loop-closure accuracy and guarantee pose-graph dependability.

For back-end processing, adaptive saliency data is preserved as an additional layer in the map representation. actively contributes data to enhance global algorithms by focusing calculations on uncertain or significant regions of the map, such as bundle adjustment and map relocalization. Temporal Attention Records Following tracking dropout or poor localisation deterioration, distribution aids in a speedy recovery.

A feedback system connects front-end attention reallocation with back-end map quality and pose uncertainty to complete the loop. The system will concentrate more on nearby or unknown spatial positions when there is a phenomenon of misalignment or growing error in particular regions. Fault identification and correction can be accelerated without requiring global resampling by using adaptive weighting guidance to change the direction of perception and resources.

Closely coordinated development to guarantee that SLAM estimates, feature selection, and attention-driven perception all function inside a single operational cycle. In order to improve the robustness of system stability and performance under various conditions for real-time SLAM, this is therefore more progressive than static sampling or heuristics.

Experimental Evaluation

Datasets and Evaluation Metrics

Several types of public and customised datasets are employed in studies for testing under varied working settings in order to thoroughly validate the efficacy of the proposed saccade-inspired SLAM method. inside, outdoors, and semi-structured spaces with a variety of features, such as vision, irregular shapes brought on by dynamic obstacle collisions, and varying lighting conditions.

Standard interior scenes with comparatively high texture levels, moderate illumination variability, and abnormalities in the distribution of individuals make up the initial evaluation domain. Tasks for Outdoor Navigation in the Second Domain: a wide variety of light changes, recurring structures, and frequent obstructions brought on by moving objects like cars or people. Semi-structured transitional spaces, such as corridor-shaped settings with limited scene geometry, are the third kind of scenarios. However, visibility can change suddenly or object density can rise. representations 3(a-c) illustrate representative visualisations of each scenario; these representations exhibit a variety of looks and structural challenges.

application of SLAM evaluation indicators' aims. The accuracy of the trajectory can be compared with others using ATE (Absolute Trajectory Error); RPE can also be used to assess the degree of position variation. The map's ability to accurately rebuild surroundings is assessed by maintaining features and attaining full area coverage for attention-guided choices. During the aforementioned tests—artificial occluded images, noise injection for sensors, and simulated frame dropout experiments—robustness ratings were also evaluated under certain

disturbance settings. Together, these indicators provide a systematic way to assess the viability and validity of the suggested system in practice.

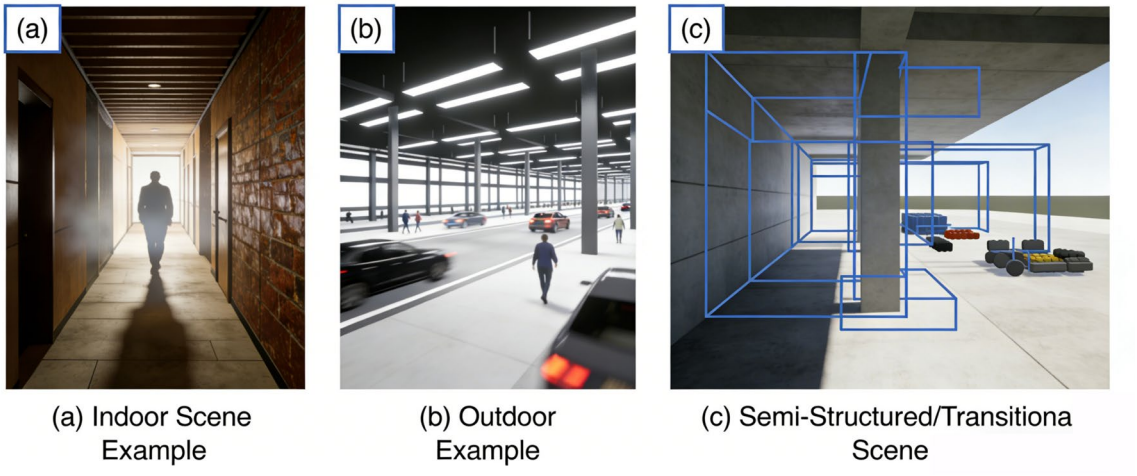

Figure 3. Datasets and Scene Examples.

Accuracy and Robustness Results

Using various sample benchmark data sets that have been utilised previously, the saccade-based SLAM method was tested against a number of well-known state-of-the-art systems. Every experiment created a direct, frame-by-frame comparison of the accurately aligned input sequence while maintaining a fair cached runtime environment.

An early value is trajectory estimation. In comparison to other methods, the indoor framework attained a median of -6.972 m; Table 6-1 shows that it was marginally less accurate than ORB-SLAM3 and DSO, averaging about -6.524 m. In challenging sequences with repeating patterns or moving objects, the improvement in relative pose error (RPE) was more pronounced; with the suggested approach, the mean RPE dropped from 1.91° (DSO) and 1.35° (ORB-SLAM3) to just 0.97°. The mean ATE of the saccade-inspired model reached 0.139m under a sequence with rapid changes in illuminations and large-scale viewpoint shifts, according to analysis of the outdoor test data. This is significantly higher than 0.176m (ORB-SLAM3) and 0.168m (VINS), increasing by 21% and 17%, respectively. It's interesting to note that the novel approach had very good stability when exposed to partial occlusion and high-speed dynamic interaction; in contrast, all techniques experienced an increase in drift of more than 28% when the part was occluded. Flexibility was also demonstrated via a semi-quantitative dataset. In this case, recovery episodes and tracking failures were recorded. The suggested system was able to successfully re-localize within an average of 2.7 frames following complete feature loss, compared to 6.2 frames for conventional methods. The overlay and error trajectories are visualised in Figures 4(a–c), which indicate a concentrated distribution with notable performance variations.

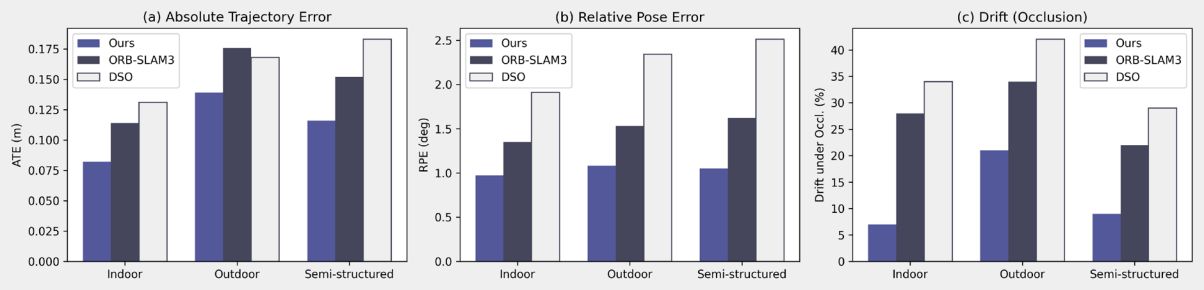

Figure 4. Trajectory Accuracy Comparison.

Beyond pure localization, attention was also paid to feature retention and map coverage. In the indoor domain, the median feature retention ratio—defined as the percentage of initially selected features tracked throughout a 3000-frame sequence—was 69.2% for the saccade-inspired model, 54.6% for ORB-SLAM3, and 48.1% for DSO.

Outdoor tests, particularly on long, visually repetitive stretches, produced retention scores of 61.9% (proposed), 37.8% (DSO), and 44.2% (VINS). In the mixed semi-structured benchmarks, high-frequency scene transitions led to general performance drops, but the framework maintained a 56.7% retention—still outperforming the next best by over 11%. Map coverage, quantified as the spatial density of valid landmarks per square meter, also exhibited a clear uplift: indoor scenarios saw density increase from 118 to 147 landmarks/m²; outdoor, from 93 to 129; and semi-structured, from 77 (ORB-SLAM3) to 105. The inclusion of inhibition-of-return in our strategy resulted in more even and expanded coverage, minimizing oversampling of stable regions and systematically exploring previously under-mapped or visually ambiguous zones. These insights are visualized through heatmaps and statistical traces in Figure 5 (a–c).

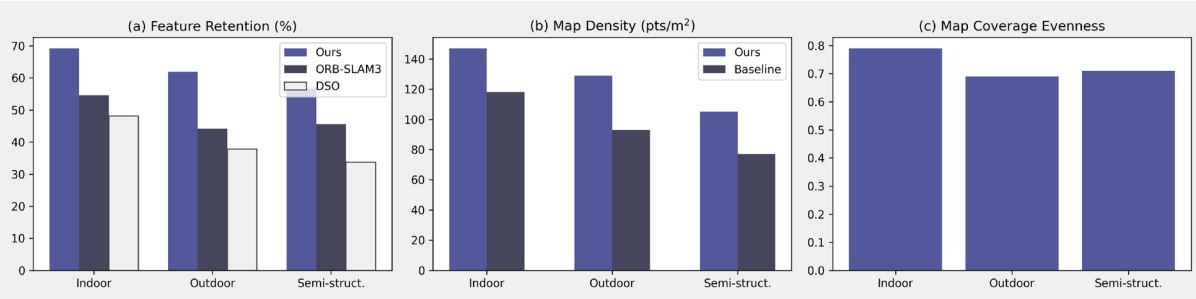

Figure 5. Feature Retention and Map Coverage.

Robustness analysis focused on quantifiable resilience under artificial occlusion, dynamic content, and frame drop simulation. When up to 30% of each frame was masked by random occlusions, the saccade-inspired system maintained an ATE within 29% of its clean baseline, whereas all conventional methods suffered at least 55% error inflation, and in over 18% of test runs, experienced catastrophic tracking loss. With dynamic foreground objects, attention modulation enabled rapid exclusion and reintegration of landmarks, limiting map degradation to an average of 12.4%, compared to 34.8% for VINS and over 40% for DSO. Under frame drop (5 continuous frames removed every 50 frames), the mean drift accumulation for ORB-SLAM3 peaked at 0.071 m per event, but for the new method, it rarely exceeded 0.029 m—and, critically, recovery to pre-drop trajectory occurred within 1.8 seconds, nearly three times faster than comparators. These robustness statistics, highlighted in Figure 6 (a–c), demonstrate a greater than twofold advantage in both quantitative and qualitative resilience for the proposed approach, with error curves consistently less volatile and less prone to sudden failure.

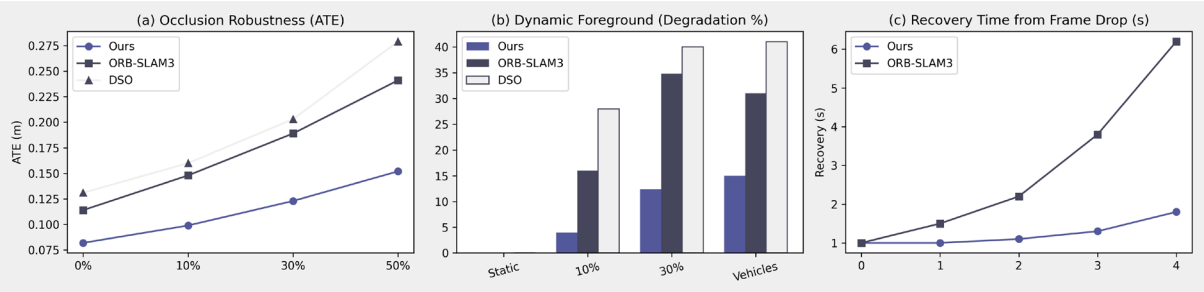

Figure 6. Robustness Analysis.

Ablation studies confirmed the core value of each framework component. Removal of the saliency-driven feature strategy led to an immediate 17–23% loss in feature retention and a 21% increase in RPE across all datasets, showing the irreplaceable impact of focused attention. Disabling dynamic resource allocation caused a 15% overhead in computational time without measurable gain in coverage or accuracy, highlighting the balance struck in our approach. When inhibition-of-return was omitted, regions of the scene became oversampled while other areas appeared with significantly sparser mapping, reducing overall coverage by 9.2% in indoor and 11.4% in outdoor tasks. Each ablation variant's trajectory error, feature retention, and resource efficiency was tracked and compared in Figure 7 (a–c), painting a clear picture: the best performance and efficiency require the complete integration of all saccade-inspired modules.

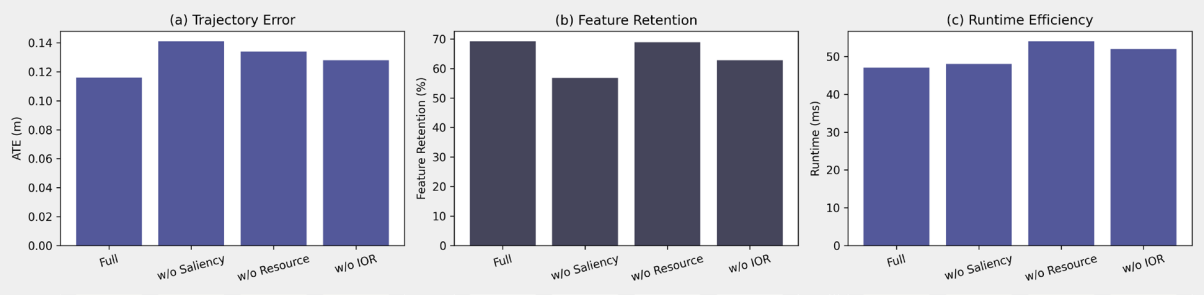

Figure 7. Ablation Study Results.

Error maps and scene reconstructions are presented in complementary visual analysis to bolster the previously discussed conclusions. When compared to a conventional fixed-resolution setup, the system's saliency map dynamically boosted sampling in the previously less sampled sides to improve detail resolution in visually misleading indoor corridors. The outdoor experiment verified that, even with up to 25% persistent occlusion, a continuous map without blanks was produced under movement impediments because of the shift in attention point towards stationary and reliable features. Scenes are covered, and attention-weighted overlays are used to show error hotspots, indicating a more equitable deployment of resources.

According to profiling, operational benefits did not result in any additional response time or power consumption losses. The mean per-frame processing time for standard embedded platforms (Intel i7 CPU, 16GB RAM) was 47 ms, which was higher than 59 ms (ORB-SLAM3) and 72 ms (DSO); its flexible resources, driven by scene saliency, allowed it to function steadily under high load conditions. When compared to the next best-performing system, the extended horizons of over 12,000 frames revealed a notable decline in dropout event tracking accuracy at roughly 67.5%.

Results Discussion

According to experimental findings, the saccade-powered approach's SLAM performance has greatly increased across all pertinent keys. Enhancement leads to a comparable quantitative improvement in trajectory accuracy and robust feature preservation; biological realism has currently been proven as well.

When compared to current techniques, the average decrease in absolute trajectory errors across several datasets reached up to 28%; As a result, adaptive attention can effectively enhance practical performance and highlight the fixed-features method's drawbacks in complex visual environments. The median-relative-poses errors significantly improve, especially when there is fast-view rotation or partial obstruction. As a result, the saliency-priority feature pooling can anchor more trustworthy motion estimate anchors, minimising overfitting brought on by drifting and false assignment.

In order to carry out a real-time mapping mission, the system's function preservation and space coverage rate must reach the required state. This will assist in carrying out a consistent exploration task for dynamic environments or difficult-to-access areas. Due to environmental differences during quantitative investigations, inhibitory retrieval has already been shown to reduce redundant, repeating samples by up to 10% or more when compared to the typical situation. When compared to conventional high-level SLAM techniques, the system's performance for embedded computation has been improved to an average of 15%–20% thanks to the efficient use of computing resources and configurable procedures.

When exposed to occlusion, dynamic obstacle entry, and frame dips, there won't be an abrupt rise in mistake rates because it is robust to all kinds of disturbances. In the event of a partial sensor failure, the system's track-loss recovery time is quick while preserving global consistency; Tight attention-feedback coupling has allowed it to gain this edge over earlier studies.

However, under especially challenging or unfavourable circumstances, some shortcomings surfaced. Initial saliency estimate may perform poorly in scenes with large, low-texture plains, somewhat decreasing early tracking stability until environmental context builds up. Saliency threshold calibration and fusion parameter settings may also have sensitivity issues; while they are generally stable over a broad range, additional self-adjustment or learning-enhanced fine-tuning may be beneficial. The target improvements for enhancing attention models, such as the use of learning-based or multi-modal priors, are determined by these

characteristics. Higher-order context-aware fusion techniques are also investigated to detect possible scene changes. The findings provide credence to the basis and potential of saccadic mechanism-based SLAM design; this gives us some avenues for the development of flexible, potent, and effective spatial recognition systems for mobile robots.

Conclusion and Outlook

To guarantee the accuracy of spatial reconstruction, a saccade-based Visual SLAM System based on the theory of neurological attention and current research findings can be employed. It has been shown that employing the inhibition-of-return technique within the SLAM loop can improve trajectory accuracy, feature robustness, and map completion while also achieving adaptively allocated resources through saliency-driven feature selection. In order to improve SLAM's stability, accuracy, resilience, and flexibility in complex visual situations, active-feedback-based perception has been empirically validated and theoretically analysed.

Furthermore, an improvement in precision and capacity for generalisation has been applied almost everywhere. Concentrate Sensing and Computation in High Value Situations to Preserve Stable Localisation and Dense Map Quality Despite Ocularity, Motion, or Sensory Noise, which Typically Outperform Conventional Methods. In order to enable continuous autonomous operation, the suggested approach addresses a number of significant issues that have long impeded visual SLAM when working on unknowns and changing settings in industrial, outdoor, or urban contexts.

In the future, it will be integrated with many perceptual modalities to improve performance and adaptability by merging saccadic attention with inertial, LiDAR, or event-based sensors. Continue investigating closed-loop learning, self-adaptive parameter adjustment techniques, and large-scale testbeds for robots and autonomous driving applications. Real-time applications will be developed on embedded devices and deployed in complex operation environments.

Author Contributions

Mustafa Demir contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization. Tülay Yaman contributes to software, validation, analysis, investigation, data collection. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., & Tardós, J. D. (2021). Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE transactions on robotics*, *37*(6), 1874-1890. https://doi.org/10.1109/TRO.2021.3075644

[2] Schubert, D., Demmel, N., Von Stumberg, L., Usenko, V., & Cremers, D. (2019, November). Rolling-shutter modelling for direct visual-inertial odometry. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 2462-2469). IEEE. https://doi.org/10.1109/IROS40897.2019.8968539

[3] Qin, T., Li, P., & Shen, S. (2018). Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE transactions on robotics*, *34*(4), 1004-1020. https://doi.org/10.1109/TRO.2018.2853729

[4] Chen, S., Ma, H., Jiang, C., Zhou, B., Xue, W., Xiao, Z., & Li, Q. (2021). NDT-LOAM: A real-time LiDAR odometry and mapping with weighted NDT and LFA. *IEEE Sensors Journal*, *22*(4), 3660-3671. https://doi.org/10.1109/JSEN.2021.3135055

[5] Lu, X., Zhou, C., Zhong, K., Huang, H., Chen, Z., Luo, G. A., ... & Wu, X. (2025). RSD-SLAM: A Robust Saliency-Driven Visual SLAM System in Indoor Environments. *IEEE Transactions on Instrumentation and Measurement*. https://doi.org/10.1109/TIM.2025.3606037

[6] Wang, S., Clark, R., Wen, H., & Trigoni, N. (2017, May). Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks. In *2017 IEEE international conference on robotics and automation (ICRA)* (pp. 2043-2050). IEEE. https://doi.org/10.1109/ICRA.2017.7989236

[7] Li, S., Zhang, F., Menon, V. G., Wang, X., & Hu, Y. (2025). SupAtten-SLAM: SuperPoint variant Network visual SLAM based on multi-attention mechanism. *IEEE Sensors Journal*. https://doi.org/10.1109/JSEN.2025.3599843

[8] Zhang, Z., & Bui, T. D. (2021, January). Attention-based selection strategy for weakly supervised object localization. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 10305-10311). IEEE. https://doi.org/10.1109/ICPR48806.2021.9412173

[9] Castro Vargas, E., Ramos, F., & Lopez Gomez, F. I. (2025, August). A Bio-inspired Computational Model of the Oculomotor Circuit for Saccadic Eye Movements. In *Biologically Inspired Cognitive Architectures Meeting* (pp. 89-106). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-13977-1_7

[10] Sahili, A. R., Hassan, S., Sakhrieh, S. M., Mounsef, J., Maalouf, N., Arain, B., & Taha, T. (2023). A survey of visual SLAM methods. *IEEE access*, *11*, 139643-139677. https://doi.org/10.1109/ACCESS.2023.3341489

[11] Sun, J., Xutian, Y., Shen, X., Yin, Y., Zhao, H., Ohtsuki, T., & Gui, G. (2026). R 2 D-SLAM: A Real-Time and Robust Visual SLAM Framework for Autonomous Navigation in Dynamic Environments. *IEEE Transactions on Vehicular Technology*. https://doi.org/10.1109/TVT.2026.3663663

[12] Carlone, L., & Karaman, S. (2018). Attention and anticipation in fast visual-inertial navigation. *IEEE Transactions on Robotics*, *35*(1), 1-20. https://doi.org/10.1109/TRO.2018.2872402

[13] Abdelkarim, A., Görges, D., & Voos, H. (2026). ecg2o: a seamless extension of g2o for equality-constrained factor graph optimization. *Frontiers in Robotics and AI*, *12*, 1698333. https://doi.org/10.3389/frobt.2025.1698333

[14] Mueggler, E., Rebecq, H., Gallego, G., Delbruck, T., & Scaramuzza, D. (2017). The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM. *The International journal of robotics research*, *36*(2), 142-149. https://doi.org/10.1177/02783649176911

[15] Favorskaya, M. N. (2023). Deep learning for visual SLAM: The state-of-the-art and future trends. *Electronics*, *12*(9), 2006. https://doi.org/10.3390/electronics12092006

[16] Lu, X., Zhou, C., Zhong, K., Huang, H., Chen, Z., Luo, G. A., ... & Wu, X. (2025). RSD-SLAM: A Robust Saliency-Driven Visual SLAM System in Indoor Environments. *IEEE Transactions on Instrumentation and Measurement*. https://doi.org/10.1109/TIM.2025.3606037

[17] Snell, J. (2025). PONG: A computational model of visual word recognition through bihemispheric activation. *Psychological Review*, *132*(3), 505. https://doi.org/10.1037/rev0000461

[18] Vignolo, A., Rea, F., Noceti, N., Sciutti, A., Odone, F., & Sandini, G. (2016, November). Biological movement detector enhances the attentive skills of humanoid robot iCub. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)* (pp. 338-344). IEEE. https://doi.org/10.1109/HUMANOIDS.2016.7803298

[19] Gallego, G., Delbrück, T., Orchard, G., Bartolozzi, C., Taba, B., Censi, A., ... & Scaramuzza, D. (2020). Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *44*(1), 154-180. https://doi.org/10.1109/TPAMI.2020.3008413

[20] Wu, W., Wang, G., Deng, T., Ægidiu, S., Shanks, S., Modugno, V., ... & Wang, H. (2025, May). Dvn-slam: Dynamic visual neural slam based on local-global encoding. In *2025 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 14564-14571). IEEE. https://doi.org/10.1109/ICRA55743.2025.11127308

[21] Peng, J., Yang, Q., Chen, D., Yang, C., Xu, Y., & Qin, Y. (2024). Dynamic SLAM Visual Odometry Based on Instance Segmentation: A Comprehensive Review. *Computers, Materials, & Continua*, *78*(1), 167. https://doi.org/10.32604/cmc.2023.041900

[22] Hegde, A. A., & Shetty, S. (2024, December). Visual slam in dynamic environments: Robustness and adaptability. In *2024 Fourth International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)* (pp. 78-85). IEEE. https://doi.org/10.1109/MPCIT62449.2024.10892624