Map consistency and long-term localization in monocular SLAM

A.L.Sulamidinov
Jizzakh State Pedagogical University

Abstract: *The promise of monocular Simultaneous Localization and Mapping (SLAM) is profound: endowing autonomous systems with the ability to perceive, understand, and navigate uncharted environments using a single camera, the most ubiquitous and biologically-inspired sensor. For over two decades, the core mathematical and algorithmic challenges of real-time structure from motion have been largely conquered, enabling robust operation in constrained, short-term sequences. However, the leap from a functioning demonstrator to a reliable, persistent spatial intelligence agent hinges on solving two intertwined, grand challenges: long-term localization and global map consistency. This article contends that these are not merely peripheral issues but the central bottlenecks preventing the deployment of monocular SLAM in real-world applications that demand longevity, such as domestic robotics, augmented reality, and autonomous inspection. We explore the fundamental limitations of purely visual methods in scale-drift management, the catastrophic impact of incremental error accumulation on map utility, and the necessity of map maintenance and correction over extended periods. The discourse will navigate through the evolution of techniques from covisibility graphs and pose-graph optimization to modern explicit and implicit mapping strategies, analyzing how each confronts the demons of long-term operation. Ultimately, we argue that the future of robust, consistent monocular SLAM lies not in increasingly complex purely visual systems, but in the principled, tight integration of other weak or intermittent cues - be it inertial measurements, sparse depth, semantic permanence, or learned priors - to anchor the visual SLAM system to a stable, consistent, and reusable representation of the world.*
Keywords: *monocular SLAM, long-term autonomy, map consistency, visual-inertial odometry, loop closure detection, neural implicit representations*

Introduction

Monocular SLAM stands as a cornerstone of modern robotics and computer vision, a discipline dedicated to solving the chicken-and-egg problem of concurrently estimating the trajectory of a moving camera and the three-dimensional structure of its surroundings. Its allure is rooted in the simplicity, low cost, and rich information density of the imaging sensor. From the seminal breakthroughs of parallel tracking and mapping (PTAM) to the current era of direct, indirect, and hybrid methods, the community has achieved remarkable feats in real-time performance, robustness to motion blur, and handling of low-texture environments. Yet, a vast chasm separates a system that works for a few minutes in a laboratory from one that can operate for days, weeks, or months in a dynamic, evolving world, returning to a known place with centimeter-level accuracy after arbitrary absences.

The core of this chasm is the issue of consistency. In a probabilistic sense, a consistent map is one where the estimated state (poses and landmarks) converges to the true state, with uncertainty bounds that accurately reflect the real error. In a more pragmatic, engineering sense, a consistent map is one that is globally coherent, free from cumulative drift, and can be reliably used for re-localization and path planning at any future time. Monocular vision, for all its strengths, introduces two fundamental infirmities that sabotage consistency: the lack of absolute scale observability and the incremental nature of bundle adjustment. Every frame-to-frame or local bundle adjustment, while optimizing local geometry, introduces minute errors in rotation, translation, and point positions. These errors, particularly in translation direction and scale, accumulate unbounded over long trajectories. The resulting map, while locally accurate, becomes a distorted, rubber-sheeted version of reality - a world

where loops may close correctly in the visual feature space but create impossible geometric distortions, and where a robot cannot distinguish between a position it visited an hour ago and one visited a week ago.

Therefore, the twin pillars of this discussion - long-term localization and map consistency - are intrinsically linked. Long-term localization (the ability to precisely determine one's position within a previously built map, regardless of changes in appearance, lighting, or partial occlusion) is impossible without a consistent map. Conversely, a map cannot be maintained as consistent without mechanisms for detecting and correcting drift, which inherently requires the ability to re-localize within it from drastically different viewpoints and under changed conditions. This article will dissect this symbiotic relationship, examining the historical and contemporary approaches to taming the inconsistencies inherent in monocular SLAM.

The Roots of Inconsistency: Scale Drift and Error Accumulation

To understand the solutions, one must first appreciate the profundity of the problem. A monocular camera is a projective sensor; it captures bearing information but not metric distance. Depth is triangulated from parallax over a sequence of frames. The scale of the reconstructed world is set arbitrarily, often from an initialization procedure. While this scale can be maintained relatively well locally through careful bundle adjustment, it is not observable in a global sense. Any minute bias in feature matching, camera calibration, or motion estimation propagates into the scale estimate. Over hundreds or thousands of frames, this scale can slowly expand or contract, a phenomenon known as scale drift. This drift couples with rotational and translational drift to create the characteristic bending of long trajectories often seen in monocular SLAM outputs before loop closure.

The primary tool for mitigating this incremental error is the pose-graph optimization framework. Here, the map is abstracted into a graph where nodes represent keyframe poses and edges represent spatial constraints between them. These constraints come from relative pose measurements derived from local bundle adjustment (sequential edges) and, crucially, from loop closures. When the system recognizes a previously visited location, it creates a new edge between the current keyframe and the past keyframe, providing a potent constraint that contradicts the accumulated drift. The subsequent optimization of the entire graph distributes the loop closure error back along the trajectory, rectifying the global map consistency.

However, this elegant solution belies a host of sub-problems. First, loop closure detection itself is a monumental challenge in long-term operation. Appearance-based methods using bag-of-words models must contend with extreme perceptual changes: day to night, summer to winter, and changes in furniture, people, or parked cars. A false positive loop closure - an incorrect match - is catastrophic, as it injects a fundamentally wrong constraint into the pose-graph, irreparably corrupting the map. Robust systems require geometric verification and sometimes rejection mechanisms to maintain integrity. Second, even a correct loop closure in a large map presents a computational challenge. Full bundle adjustment on all points and keyframes becomes prohibitive. Efficient pose-graph optimization, while faster, marginalizes out the point cloud, which can sometimes lead to sub-optimal solutions. The management of this optimization, deciding when to perform a local versus global adjustment, is key to maintaining real-time performance without sacrificing consistency.

Mapping Strategies: From Explicit Features to Implicit Representations

The choice of map representation fundamentally dictates strategies for maintaining consistency. Traditional feature-based SLAM systems maintain an explicit sparse map of 3D points (landmarks), each associated with one or more visual descriptors. Consistency is managed through the lifecycle of these points: their creation, triangulation, selection for optimization, and eventual culling or replacement. In long-term operation, this map becomes a living entity. The system must not only add new points but also identify and remove obsolete ones that have become unstable due to mismatches or changes in the environment. More importantly, it must retain a sufficient number of persistent

landmarks that act as anchor points across multiple sessions. This leads to concepts like “core” keyframes and “covisibility” graphs, where the map is not a monolithic collection but a structured network of interconnected frames and landmarks. Pruning this network intelligently to retain the essential skeletal structure of the environment is critical for preventing unbounded memory growth and ensuring efficient re-localization.

A significant shift in the pursuit of consistency has been the move towards dense or semi-dense mapping, often using direct methods that optimize photometric error. While these can produce more visually pleasing and navigationally useful maps, they compound the consistency problem. The photometric error landscape is highly non-convex and more susceptible to lighting changes. Maintaining a globally consistent dense map from a monocular camera, especially over large scales, remains exceptionally challenging. This has spurred the development of implicit neural representations, such as Neural Radiance Fields (NeRFs) and their variants for SLAM. These methods represent the map as a continuous function (a neural network) that encodes geometry and appearance. Their promise for consistency lies in their global nature; the network parameters are optimized against all observed data simultaneously, potentially enforcing a more globally coherent scene representation. However, they currently struggle with catastrophic forgetting when trained incrementally (the plasticity-stability dilemma), and their computational cost for optimization is high, making real-time, large-scale consistency management an open research frontier.

The Imperative of Multi-Modal Anchoring

Given the intrinsic limitations of a single, purely visual stream, the most promising path toward lifelong map consistency lies in multi-modal anchoring. The visual SLAM system must be provided with occasional, absolute references to ground its drifting estimate. This does not necessarily mean relying on expensive, high-fidelity sensors like laser scanners. Instead, it involves the strategic use of weak, sparse, or intermittent cues.

The classic partner is an Inertial Measurement Unit (IMU), giving rise to Visual-Inertial Odometry (VIO). The IMU provides proprioceptive measurements of acceleration and angular velocity, which are observables of metric scale and gravity direction. Tightly coupled VIO, where visual and inertial measurements are fused in a joint optimization, drastically reduces short-term drift and provides a locally accurate, metric scale. The IMU acts as a high-frequency stabilizer, while vision corrects the low-frequency bias drift of the IMU. For long-term consistency, however, VIO alone is insufficient; it still drifts, albeit at a much slower rate. The consistent map is now built and maintained within a visual-inertial framework, where loop closures correct the slower-growing inertial-visual drift.

Beyond inertia, other anchors are emerging. Sparse depth measurements, from a single-beam lidar or a time-of-flight sensor, provide infrequent but absolute depth values that pin the scale of the visual map. Even more abstract are semantic and structural anchors. The recognition of permanent objects (walls, doors, windows) or planar structures (floors, ceilings) allows the imposition of high-level constraints. For instance, enforcing the orthogonality of walls or the parallelism of floor and ceiling can guide optimization toward a globally consistent layout. Semantic permanence - the understanding that a building’s structure changes more slowly than the transient objects within it - can inform landmark selection and map pruning, ensuring the map’s backbone is composed of stable elements.

Perhaps the most powerful anchor for long-term operation is the ability to relocalize within a prior map under severe appearance change. This transcends traditional loop closure. It involves storing not just a map of points, but also a more sophisticated representation of place, perhaps leveraging learned features from deep networks that are invariant to lighting and weather, or creating semantic maps. When a system can reliably determine its position within a stored map from a cold start, it can then use that prior map as a fixed, consistent reference frame into which it can integrate new observations, performing continual mapping and lifelong learning without consistency loss.

Conclusion

The journey of monocular SLAM from a compelling proof-of-concept to a foundational technology for autonomous systems is a journey toward consistency. It is a shift in focus from the geometry of the moment to the persistence of place. As we have explored, the challenges are deep-rooted, stemming from the sensor itself, but they are not insurmountable. The evolution from naive sequential mapping to pose-graph optimization with robust loop closure laid the first groundwork. The ongoing refinement of map representations, from explicit sparse features to implicit neural fields, offers new avenues for encoding global coherence.

However, the analysis presented herein leads to a definitive conclusion: the monocular camera, in isolation, is insufficient for the task of building and maintaining a consistent world model over the long term. Its strength is in rich, detailed relative measurement, but it requires anchoring. The future of robust, long-term monocular SLAM - or more accurately, vision-centric SLAM - lies in its role as the primary but not sole sensor. It must be embedded within a perceptual framework that can assimilate and leverage any available source of metric or semantic anchoring: inertial dynamics, sparse depth, universal semantic priors, and the power of learned representations for place recognition. The map of the future will not be a static point cloud but a dynamic, multi-layered, and persistent spatial database, continually updated and corrected, where geometric precision is maintained not just by a chain of visual measurements, but by its firm connection to the immutable properties of the physical world. Achieving this vision is the key to unlocking the true potential of monocular SLAM for robots that share our spaces not for minutes, but for years.

References

1. Sultanov, I. R. (2025). Model predictive control of a distillation column based on a MIMO model. Academic Journal of Science, Technology and Education, 1(SI1), 41-45.
2. Toshmatov, S. T. (2025). Analysis of distortion sources in three-stage Class B audio power amplifiers. Academic Journal of Science, Technology and Education, 1(SI1), 35-40.
3. Jalilov, M. (2025). Integration of XArm robots with Tenzo force sensors for intelligent manipulation systems. Academic Journal of Science, Technology and Education, 1(SI1), 17-21.
4. Sabirov, U. K. (2025). Application of multi-criteria decision-making models in dairy product storage processes. Academic Journal of Science, Technology and Education, 1(SI1), 30-34.
5. Sokhibova, Z. M. (2025). Chemical processing of silicon semiconductor materials and their physical fundamentals. Academic Journal of Science, Technology and Education, 1(SI1), 22-25.
6. Nematov, A. (2025). Modern approaches to intelligent automation of agricultural technological processes based on artificial intelligence. Academic Journal of Science, Technology and Education, 1(SI1), 26-29.
7. Kholikov, A. S. (2025). Distance and computer-based teaching system of physics for engineering personnel working in industry. Academic Journal of Science, Technology and Education, 1(SI1), 8-11.
8. Kholmurotov, B. T. (2025). Determining the dynamics of the drying agent in the drying chamber. Academic Journal of Science, Technology and Education, 1(SI1), 55-60.
9. Kholiqova, M. K. (2025). Methodology for Implementing Project-Based Learning in Developing Independent and Creative Thinking of Students. Academic Journal of Science, Technology and Education, 1(8), 67-75.
10. Turgunbaev, R. (2025). Training dynamics in modern neural network optimization. Academic Journal of Science, Technology and Education, 1(8), 51-57.
11. Fayziyev, A. N. (2025). Clinic of immunogenetic conditions of juvenile rheumatoid arthritis in children. Academic Journal of Science, Technology and Education, 1(8), 11-15.
12. Fayziyev, A. N. (2025). Immunogenetic markers and clinical expression in juvenile rheumatoid arthritis. Academic Journal of Science, Technology and Education, 1(8), 20-24.

13. Alieva, N., Rasuleva, M., & Xalilova, S. (2025). Analysis of artificial intelligence integration in modern learning systems. Academic Journal of Science, Technology and Education, 1(8), 83-86.
14. Inogamova, N. (2025). The use of artificial intelligence for positive pedagogical purposes in teaching french as a foreign language to students of international law and international economic relations. Academic Journal of Science, Technology and Education, 1(8), 25-27.
15. Djuraeva, Z. (2025). Cultural and lexical features of gastronomic metaphors. Academic Journal of Science, Technology and Education, 1(8), 76-79.
16. Jurayeva, S. (2025). Fashion design as a self-branding strategy: methods of personal image building by media influencers. Academic Journal of Science, Technology and Education, 1(8), 3-5.
17. Ruziokhunov, D., & Yuldasheva, U. (2025). The role of government tax regulations on attracting foreign investments. Academic Journal of Science, Technology and Education, 1(8), 6-10.
18. Djuraeva, Z. (2025). Rendering the names of dishes in French and Uzbek and their comparative analysis. Academic Journal of Science, Technology and Education, 1(8), 80-82.
19. Ashurova, S. B. (2025). The importance of international educational programs in the development of vocational education. Academic Journal of Science, Technology and Education, 1(8), 58-60.
20. Rajabov, S. B., & ogli Alimov, J. S. (2025). Modern Approaches to Modernizing the Management System in Higher Education through Digital Technologies. Academic Journal of Science, Technology and Education, 1(5), 34-38.
21. Abdurahmonov, H. (2025). Logarithmic functions. Academic Journal of Science, Technology and Education, 1(5), 50-51.
22. Abduxalimova, N. X., Sunnatov, D. H., Alikhodjaev, S. S., Toirova, A. A., Tirkasheva, D. D., Shamsieva, O. B., & Xatamov, U. A. (2025). Comparative evaluation of traditional and digital methods for recording centric occlusion in prosthodontics. Academic Journal of Science, Technology and Education, 1(5), 29-33.
23. Asatullayev, R. B., & Latifova, S. N. (2025). Prescriptions and drugs. Academic Journal of Science, Technology and Education, 1(7), 50-52.
24. Nuriddinova, G. (2026). CULTIVATING THE ENGAGED READER IN BOOKS AND LIBRARIES. European Review of Contemporary Arts and Humanities, 2(1), 17-24.
25. Madaminov, N. (2026). THE ROLE OF THE ENSEMBLE IN SOLO PERFORMANCE. European Review of Contemporary Arts and Humanities, 2(1), 63-66.
26. Khaydarova, M. K. (2026). THE CONCEPT OF AESTHETIC THINKING IN VISUAL ARTS AND ITS PEDAGOGICAL SIGNIFICANCE. European Review of Contemporary Arts and Humanities, 2(1), 45-47.
27. Xalilov, T. (2026). A MODEL FOR DEVELOPING ETHICAL-MORAL COMPETENCIES IN GENERAL SECONDARY EDUCATION STUDENTS THROUGH THE MEANS OF NATIONAL MUSIC. European Review of Contemporary Arts and Humanities, 2(1), 71-75.
28. Muminov, M. (2026). BUILDING COLLABORATIVE SKILLS FOR PIANISTS IN UNDERGRADUATE PROGRAMS. European Review of Contemporary Arts and Humanities, 2(1), 36-41.
29. Yunusov, G. O. (2026). PATTERNS OF IDENTITY IN UZBEK FOLK MUSIC ARTS. European Review of Contemporary Arts and Humanities, 2(1), 25-29.
30. Fayziyev, A. N. (2026). CLINICAL LANDSCAPE SHAPED BY IMMUNOGENETICS IN JUVENILE RHEUMATOID ARTHRITIS. European Review of Contemporary Arts and Humanities, 2(1), 11-16.
31. Sodiqov, M. (2026). FROM TECHNIQUE TO TRADITION TEACHING THE UZBEK DUTAR IN HIGHER EDUCATION. European Review of Contemporary Arts and Humanities, 2(1), 48-52.

32. Obidova, R., & Ismoilova, M. (2026). THE DUTAR'S VOICE AS MELODY AND METAPHOR IN UZBEK MUSICAL THOUGHT. European Review of Contemporary Arts and Humanities, 2(1), 53-57.
33. Ismailova, M. (2026). HAND MOTOR SKILLS AND NEUROMUSCULAR MECHANISMS IN MUSICAL PERFORMANCE: THE CASE OF DUTOR PLAYING. European Review of Contemporary Arts and Humanities, 2(1), 58-62.
34. Yuldashev, A. (2026). THE PLACE OF USING AGOGIC ELEMENTS IN INSTRUMENTAL PERFORMANCE. European Review of Contemporary Arts and Humanities, 2(1), 7-10.
35. Muydinov, F. (2026). THE IMPORTANCE OF RITUALS IN THE DEVELOPMENT OF INSTRUMENTAL MUSIC. European Review of Contemporary Arts and Humanities, 2(1), 30-31.
36. Egamberdieva, N. I. (2026). TEACHING MAQOM PRINCIPLES THROUGH THE UZBEK DUTAR REPERTOIRE. European Review of Contemporary Arts and Humanities, 2(1), 67-70.
37. Talaboyev, A. (2026). CONTEMPORARY METHODS FOR TEACHING CLASSICAL UZBEK VOCAL REPERTOIRE. European Review of Contemporary Arts and Humanities, 2(1), 3-6.
38. Turgunbaev, R. (2025). CORE DATABASE COMPETENCIES FOR LIBRARY STUDENTS. European Review of Contemporary Arts and Humanities, 1(5), 104-111.
39. Isaqov, S. S. (2025). NONVERBAL COMMUNICATION AND COHESION IN THE ORCHESTRAL ENSEMBLE. European Review of Contemporary Arts and Humanities, 1(5), 112-115.
40. Xakimxo'ja qizi Odilova, M. (2025). MODERN APPROACH AND IMPORTANCE OF DRAWING TEACHING. European Review of Contemporary Arts and Humanities, 1(5), 83-86.
41. Sattarov, F. I. (2025). THE FORMATION OF COLORS IN NATURE AND THEIR CHARACTERISTICS OF CHANGE. European Review of Contemporary Arts and Humanities, 1(5), 70-73.
42. Turg'unboy qizi Karimjonova, D. (2025). FINE ART OF UZBEKISTAN. European Review of Contemporary Arts and Humanities, 1(5), 29-33.
43. Ibragimov, S. (2025). TEACHING ACADEMIC WRITING WITH CHATGPT: OPPORTUNITIES AND LIMITATIONS. European Review of Contemporary Arts and Humanities, 1(5), 92-96.
44. Narmatova, A. P. (2025). DEVELOPING COUNTRY-SPECIFIC CULTURAL COMPETENCE IN GERMAN LANGUAGE LEARNERS THROUGH LITERARY TEXTS. European Review of Contemporary Arts and Humanities, 1(5), 97-103.
45. Akhrorova, M. (2025). THE CONCEPT OF TRAGIC FATE IN COMPARATIVE LITERATURE: EASTERN AND WESTERN PERSPECTIVES. European Review of Contemporary Arts and Humanities, 1(5), 46-50.