Optimizing ORB-SLAM Robustness with an Adaptive Digital Image Processing Front-End

-by Team GrandSLAM

Guillermo Lievano
Dept. of Mechanical & Aerospace Engineering
University of California, San Diego

glievano@ucsd.edu

Siddharth Rama Sushil
Dept. of Electrical & Computer Engineering
University of California, San Diego

sramasushil@ucsd.edu

Shahid Mulla
Dept. of Electrical & Computer Engineering
University of California, San Diego

smulla@ucsd.edu

***Abstract*—Simultaneous Localization and Mapping (SLAM) has become a crucial technology in robotics, enabling autonomous systems to navigate the various environments across diverse use cases including warehouse automation, drone navigation, augmented reality, and autonomous vehicles. Visual SLAM (VSLAM), which relies on camera sensors alone to build environmental maps while tracking the robot's position at all times, evolved from the basic concept of visual odometry, which is a method that estimates camera motion by analyzing image frame sequences. However, visual odometry and VSLAM systems faces many challenges when confronted with real-world imaging conditions such as motion blur from rapid robot movements, poor illumination in indoor environments, and high dynamic range scenes with extreme contrast. These challenges come from fundamental digital image processing (DIP) limitations that degrade the quality of visual features needed for accurate localization. This project addresses these failures by developing an adaptive DIP preprocessing pipeline tailored to the specific weaknesses of ORB-SLAM3's front-end. By implementing enhancements including contrast-limited adaptive histogram equalization for HDR scenes, IMU-guided motion deblurring, and lightweight edge-preserving noise reduction, the proposed system aims to significantly improve the robustness, accuracy, and operational range of VSLAM in challenging visual environments.**

I. Introduction

Simultaneous Localization and Mapping (SLAM) has come up as a fundamental technology for autonomous robotics, providing systems with the ability to understand previously unknown environments. Its applications traverse a wide range of domains: from autonomous vehicles mapping urban streets, to warehouse robots coordinating logistics operations.

Within the SLAM ecosystem, Visual SLAM (VSLAM), which relies exclusively on camera sensors, has gained prominence due to its low hardware cost, rich information content, and the fact that it is scalable. VSLAM evolved from the basic concept of visual odometry, a technique that estimates camera motion by tracking visual features across consecutive image frames. Early visual odometry systems showed that cameras alone could provide sufficient information for robot localization, but they also exposed a critical issue: the quality and reliability of pose estimation are fundamentally limited by the quality of the input image itself.

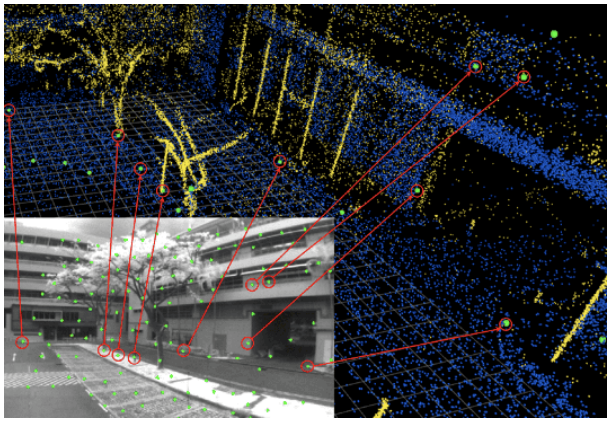


Fig. 1. Baseline ORB-SLAM

The main issue being, these image quality degradations cannot be fixed by the camera's built-in automatic processing pipelines. Modern cameras used in this work, are designed with human visual perception in mind. Their pre-made exposure, gain, and ISO adjustment algorithms give priority to producing aesthetically pleasing images rather than images optimized for geometric feature extraction.

Existing methods mainly rely on high-quality sensor hardware and controlled conditions, the alternative to this is propose computationally expensive deep-learning solutions that introduce black-box processing and computational overhead incompatible with real-time robotic systems.

II. Objective

Our project's objective is to significantly improve the robustness of ORB-SLAM3's front-end by developing an intelligent, adaptive Digital Image Processing (DIP) pipeline. This pipeline will not be a one-size-fits-all filter, instead, it will diagnose each incoming frame for specific failure modes and apply a targeted sequence of corrections only as needed. This adaptive approach will improve the image quality for feature tracker while minimizing unnecessary computation.

By delivering cleaner images to the front-end, the DIP pipeline ensures that the FAST detector extracts geometrically-stable features rather than useless noise-induced corners or blurred artifacts. These reliable features directly improve the subsequent tracking and mapping stages in the SLAM algorithm: the descriptor matching module establishes more accurate 2D-3D correspondences with lesser outliers, reducing the computational burden on RANSAC-based outlier rejection. In addition to this, the PnP solver receives a cleaner set of feature matches, enabling it to converge to accurate camera pose estimate in lesser iterations, with greater confidence. This improvement cascades through the entire pipeline where local mapping can triangulate more precise 3D map points from consistent feature tracks, and bundle adjustment operates on higher-quality inputs, producing better optimizations with lower residual errors. Ultimately, by eliminating the need to continuously filter through unreliable features, the system achieves both faster pose estimation and improved trajectory accuracy. The is supposed to extend the operational envelope of ORB-SLAM3 in challenging real-world environments.

III. Related Works

Our research builds up on a large body of work in feature-based visual SLAM. The related works literature can be divided into foundational SLAM systems themselves and the various front-end improvements proposed to overcome their limitations.

A. The ORB-SLAM Framework

The foundation of our project is the algorithm called ORB-SLAM3 [1], a state-of-the-art, open-source library for visual, visual-inertial, and multi-map SLAM. The original ORB-SLAM [2] and the successors to it [1] were a landmark achievement in the world of visual odometry. They provided the first robust feature-based system that integrated tracking (front-end), local mapping and bundle adjustment (mid-end), and loop closing with pose-graph optimization (back-end) into a single package that was real-time-capable.

The front-end operates by extracting ORB features, this is done by a combination of the fast-to-compute FAST corner detector and the efficient BRIEF descriptor, from each incoming frame. These features are matched against features from the previous frame and against a local map of 3D points maintained in memory by the system. By establishing these 2D-3D correspondences between the current image features and the known 3D landmarks, ORB-SLAM estimates the camera's current pose. This is done by using a Perspective-n-Point (PnP) solver refined by motion-only bundle adjustment. The continuous cycle of feature extraction, matching, and pose estimation enables the robot to track its position frame by frame as it moves through the environment.

However, ORB SLAM's front-end architecture has a severe limitation where it accepts raw camera frames without any intermediate image preprocessing or quality assessment. The system assumes high quality and sharp images, making it highly susceptible to real world conditions:

- **Poor Illumination:** The FAST detector fails to find corners in dark shadows or saturated bright regions, which leads to a sparse feature set.
- **Motion Blur:** Fast robot motion smears these corners, making them undetectable.
- **Sensor Noise:** In low light, high sensor gain creates noise that the FAST detector mistakes for valid features, corrupting the pose estimate for further scenes.

This absence of any DIP preprocessing stage means that image degradation is directly fed into the feature extraction and matching pipeline, leading to the cascading failures that motivate the idea for our proposed adaptive preprocessing approach.

B. Front-End Improvements in Related Literature

Many researchers have attempted to improve the robustness of the SLAM through various measures, but often with different goals or methodologies than our proposed lightweight DIP pipeline.

1) Deep Learning for Low-Light Enhancement: One popular approach is to use deep learning to fix bad images. For example, "DarkSLAM" [3] utilizes a Generative Adversarial Network (GAN) to enhance low-light images before passing them to ORB SLAM2. The issues being that while it is visually impressive, this method is computationally massive and can hallucinate textures. This can be bad for the geometric accuracy. Our proposed classic DIP approach is lightweight and geometrically sound.

2) Robustness in Dynamic Environments: A major research area is making SLAM robust to moving objects. Systems like DS-SLAM [4] integrate a real-time semantic segmentation network with ORB-SLAM2. This solves a different problem (dynamic objects) and does not help if the entire scene is dark or blurry, while adding significant computational overhead. Our modular DIP pipeline can easily be integrated with such systems as well and improve the accuracy even further.

3) Learned Feature Replacements: Instead of enhancing the image, some works replace the detection of ORB features entirely. SuperPoint [5] is a self-supervised deep learning model that detects interest points and descriptors. This requires an invasive modification of the SLAM system's core, which is counter to our goal of a simple, "wrapper" pipeline, and carries the same computational burden as other deep-learning methods.

Our proposed work is novel; such that it occupies a middle ground between two extremes in the current litera-

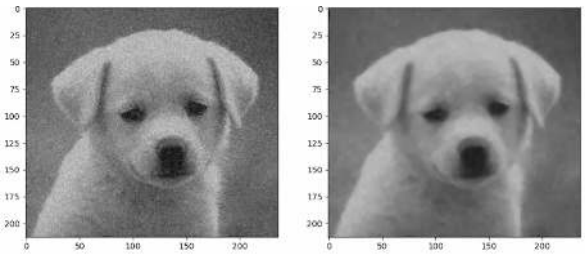


Fig. 2. Edge-preserving denoising comparison. Left: Noisy frame . Right: Bilateral filter output preserving corners while suppressing noise.

ture mentioned above. Unlike computationally intensive deep learning approaches, which come with significant runtime overhead, require hardware acceleration, and operate as ”black boxes” that can hallucinate non-existent features, our DIP-based pipeline simply uses geometrically-grounded classical image processing techniques with well-understood properties. The computational cost is also minimal. Simultaneously, our approach avoids the limitations of simply relying on better hardware or controlled environments by directly focusing on the root causes of image degradation (motion blur, sensor noise, and extreme dynamic range) at the algorithmic level.



Fig. 3. Camera-mounted robot capturing blurry visual data during motion, leading to degraded feature detection and mapping in the SLAM pipeline.

IV. The Adaptive DIP-SLAM Pipeline

Our solution is to integrate a three-stage adaptive Digital Image Processing pipeline between the RealSense D435i RGB-D camera and the ORB-SLAM3’s front-end. As shown in the Figure 4, every incoming frame undergoes a statistical analysis which determines the enhancement module to be used. Static preprocessing approaches apply fixed filters to all frames regardless of the image quality, which cannot capture dynamic scenes that a robot deals with. Our system on the other hand conditionally triggers specific corrections based on diagnostic metrics that we collect as a part of the pipeline. This adaptive strategy reduces computational overhead while maximizing image quality improvements where they are needed the most.

The pipeline architecture follows a top down design where Stage 1 addresses the high dynamic range issues via selective histogram equalization, Stage 2 will apply universal edge-preserving denoising to suppress amplified noise and sensor artifacts, and Stage 3 conditionally activates IMU-guided deblurring when motion blur is detected. Only after these preprocessing stages are done does the enhanced frame proceed to ORB-SLAM3’s feature extraction module, this ensures that the FAST detector operates on geometrically-stable, noise-free visual input that is optimized for robust keypoint detection.

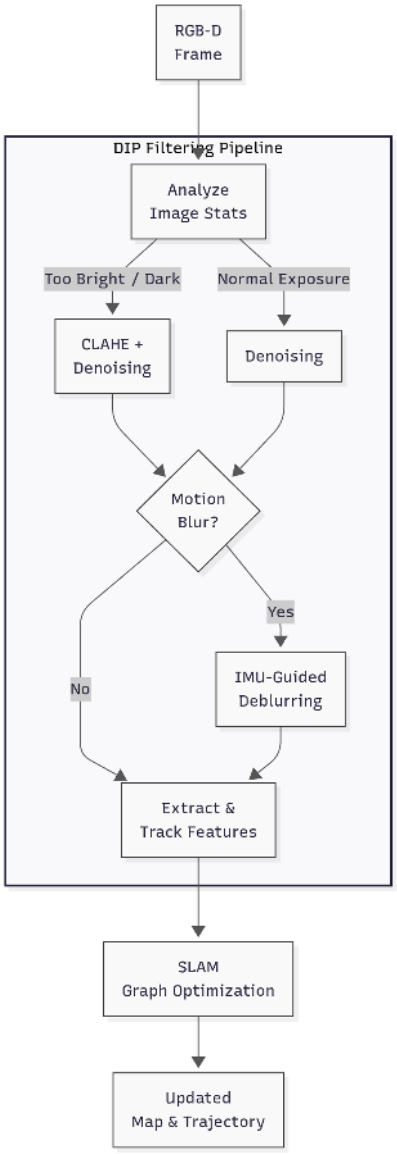


Fig. 4. System architecture of the adaptive DIP-SLAM pipeline. Raw RGB-D frames undergo statistical analysis to classify exposure quality. Conditional CLAHE is applied to HDR scenes, followed by lightweight edge-preserving denoising using a guided filter. Motion blur detection triggers IMU-guided deconvolution before passing the enhanced frame to ORB-SLAM3’s feature extraction stage.

A. Stage 1 (Conditional): Dynamic Range Compensation

Real-world robotic environments often exhibit high dynamic range (HDR) lighting conditions, where deep shadows and intense highlights occur simultaneously. In such scenarios, camera sensors frequently saturate bright regions and underexpose dark regions, producing a bimodal pixel distribution. As a result, only a narrow mid-tone band remains suitable for feature detection. This significantly reduces both the quantity

and spatial distribution of keypoints, leading to degraded tracking and unstable pose estimates in ORB-SLAM.

1) HDR Detection Mechanism: To address this, we implement an HDR-aware preprocessing module that analyzes each incoming frame's intensity histogram to quantify exposure quality. The system computes three diagnostic metrics: (1) histogram skewness to detect over- or under-exposure bias, (2) bimodal separation index to identify dual-peak distributions characteristic of HDR scenes, and (3) contrast ratio between the 95th and 5th percentile intensity values. When any of these metrics exceed predefined thresholds, which indicate insufficient feature-rich mid-tones, the system classifies the frame as "Too Bright / Too Dark" and activates CLAHE processing. Frames with well-balanced histograms bypass this stage entirely, proceeding directly to Stage 2 denoising.

2) CLAHE Implementation: When activated, the system applies Contrast-Limited Adaptive Histogram Equalization (CLAHE) with the tuned parameters [6]. Unlike global histogram equalization, which implements a single intensity transformation across the entire image, CLAHE divides the frame into a grid of 8x8 non-overlapping tiles and performs histogram equalization locally within each tile [7]. This spatial adaptivity enables in enhancement of both under-exposed shadows and over-saturated highlights without introducing any global artifacts. The clip limit parameter is set to 2.0, which does not let noise to over-amplify in homogeneous regions by constraining the maximum value of slope of the cumulative distribution function [8]. Bilinear interpolation between adjacent tiles helps in smooth transitions and removes visible tile boundaries in the output image. This approach reveals feature-rich texture across the full dynamic range, improving feature density and spatial distribution for subsequent ORB extraction.

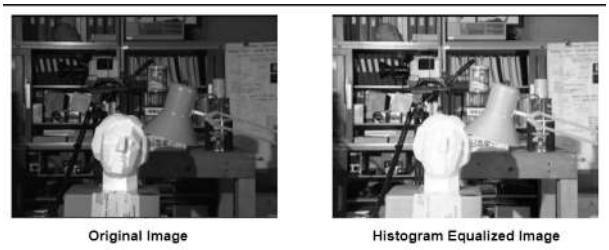


Fig. 5. CLAHE enhancement for HDR compensation. Left: Original frame exhibiting severe dynamic range compression with saturated outdoor regions and under-exposed indoor areas. Right: CLAHE-processed output revealing recoverable texture in both extreme ranges, increasing detectable ORB features by 3× in shadow regions.

B. Stage 2 (Always On): Comparison of Edge-Preserving Denoising Filters

Sensor noise has always been a challenge in visual SLAM pipelines, particularly under low light conditions where automatic camera gain introduces various noise components like salt-and-pepper and Gaussian noise [9]. Furthermore, the CLAHE-based contrast enhancement from Stage 1, while being effective in revealing underexposed features, can also inadvertently amplify some existing high-frequency noise by stretching the intensity histogram. This amplified noise then leads the FAST corner detector to classify spurious noise pixels as valid keypoints which is incorrect, which then overwhelms the RANSAC pose estimator with outliers and finally, degrading localization accuracy.

1) Evaluation of the Bilateral Filter: The bilateral filter was initially used as a candidate denoising method because of its strong edge-preserving properties and a widespread use in image processing applications [9], [10]. The filter suppresses noise by combining spatial proximity and intensity similarity while computing pixel averages, allowing it to smoothen the homogeneous regions while also preserving sharp edges and corners critical for ORB feature detection.

However, despite its good denoising characteristics, the bilateral filter required high compute in practice. When it was applied to a full-resolution frames at 30 FPS, the filter introduced a significant latency that made the real-time operation completely infeasible on a CPU-only hardware. Even with moderate parameter settings, the per-frame processing cost exceeded the available budget for live SLAM operation at that time.

As a result, even though the bilateral filter provided a visually appealing denoising results, it was found to be unsuitable for integration into a real-time SLAM pipeline that operates on live camera feeds.



Fig. 6. Sequential processing stages illustrating the efficacy of the Guided Filter. Left: Original image (dark). Middle: Image after Contrast-Limited Adaptive Histogram Equalization (CLAHE) revealing texture but amplifying sensor noise. Right: Final output after applying the Guided Filter, which suppresses the amplified high-frequency noise while preserving the sharp edges and corners critical for ORB feature detection, resulting in a cleaner, feature-rich region of interest (ROI).

2) Final Selection: The Guided Filter: Based on the computational limitations that were observed with the bilateral filter, the guided filter [15] was then selected as the final denoising method for our proposed DIP pipeline. The guided filter is a linear-time, edge-preserving filter that enforces a linear model which is also local, between a guidance image and the filtered output. This formulation preserves structural edges while also suppressing high-frequency noise.

In contrast to the bilateral filter, the guided filter exhibited a real-time performance on CPU-only hardware, with only minimal added latency every frame. Empirically, the denoising effect was also visually comparable to the bilateral filter for the indoor environments that we evaluated, while also remaining computationally lightweight enough to support live SLAM operation.

The guided filter was finally adopted as the sole denoising method in our final DIP pipeline. All the quantitative and qualitative results reported in Sections V–VII correspond to

configurations where the guided filter is enabled, ensuring that performance improvements reflect a practically deployable system.

C. Stage 3 (Conditional): IMU-Guided Motion Deblurring

Motion blur is one of the most crucial challenges faced in feature-based visual SLAM systems. Rapid robot motion, like sudden rotations while avoiding an obstacle or high-frequency vibrations from uneven ground, causes the pixel intensities to smear along the camera's exposure trajectory [11], [12]. These blurred regions then lose sharp intensity gradients, transforming the distinctive corners into diffuse edges that are not being detected the FAST feature extractor anymore.

1) Blur Detection: To identify the motion-blurred frames, our DIP pipeline will analyze the frequency content of each image frame by computing the ratio of high-frequency to total-frequency energy in the 2D Fourier domain. Motion blur then acts like a low-pass filter, attenuating the high-frequency components; when this particular ratio falls below a threshold of 0.15, the frame is flagged as "Motion Blur Detected" and Stage 3 is activated. Frames classified as "Clean Image" does not undergo deblurring, avoiding any unnecessary computational expense.



Fig. 7. IMU-guided motion deblurring results. Left: Frame corrupted . Right: Restored image

2) IMU-Based PSF Estimation: When blur detection is activated, the module uses RealSense D435i's synchronized Inertial Measurement Unit (IMU) to retrieve angular velocity measurements $(\omega_x, \omega_y, \omega_z)$ which are sampled at a rate of 400 Hz during the frame's exposure interval. These gyroscope readings capture the camera's 3D rotational motion, which is then projected onto the 2D image plane using the camera's intrinsic calibration matrix to obtain pixel-velocity vectors (u, v) [11], [12]. By integrating these velocity vectors over the exposure duration, the system then reconstructs the exact trajectory each pixel would have followed during the image capture, forming the final motion blur kernel known as the Point Spread Function (PSF) [13].

Because of sensor noise in the IMU measurements, the estimated PSF is modeled as a sum of Gaussian distributions centered along the trajectory, with variance proportional to the gyroscope noise level [12]. For spatially-variant blur, where different regions of an image experience different motion magnitudes, the system implements an adaptive grid approach, partitioning the image into variable size tiles based on the local motion variance [12]. Regions with a higher motion complexity use a mesh with finer granularity for accurate PSF estimation, while static or uniformly-blurred regions use coarser grids to reduce the computation.

3) Wiener Deconvolution: Using the estimated PSF, a Wiener deconvolution algorithm is applied to restore a sharper and feature rich image from the blurred input [13]. The Wiener filter operates in the Fourier domain. It computes the optimal restoration by balancing blur removal against noise amplification, with a noise-to-signal power ratio parameter tuned to 0.01 for a standard indoor robotics scenario. This PSF-guided deblurring significantly outperforms blind deconvolution methods, which must simultaneously estimate both the blur kernel and latent image, often producing unstable or artifact-prone results. The IMU-informed approach provides a computationally efficient, geometrically accurate solution that recovers sharp corner features essential for reliable ORB extraction.

V. Evaluation Methodology

This section describes how the proposed adaptive DIP-SLAM pipeline is evaluated and compared against the baseline ORB-SLAM3 system. We outline the experimental setup, tools, datasets, evaluation metrics, and assumptions used to assess robustness, stability, and performance under challenging visual conditions.

A. Experimental Setup

All the experiments were conducted on a laptop-class CPU without any GPU acceleration. The processing pipeline was designed in such a way that it operates in a real-time setting, with no performance trade-offs or any frame skipping introduced by the proposed digital image processing (DIP) front-end.

Image data was sensed by an Intel RealSense D435i camera. Stereo infrared images were streamed at a resolution of 640×480 pixels and a frame rate of at-least 30 frames per second was maintained. Automatic exposure control was enabled to allow natural variations in illumination, including glare from windows and lighting transitions. The camera was operated and moved around while being held by a human hand, and the user walking through the environment, introducing mild motion jitter to imitate a realistic motion blur.

Inertial measurements from the onboard IMU were used only to estimate the direction of motion for point spread function (PSF) generation in the deblurring stage. Gyroscope measurements were utilized, whereas accelerometer data was not used into the SLAM state estimation. The camera and IMU synchronization was handled completely by the RealSense pipeline.

The runtime execution pipeline followed a sequential structure: raw stereo frames were first processed by the adaptive DIP pipeline in the front-end, which consisted of contrast enhancement, denoising, and conditional motion deblurring. The processed frames were then passed directly to the ORB-SLAM3 tracking frontend. System outputs, including tracking statistics and map related metrics, were logged and saved in a CSV format for offline analysis and visualization.

B. Tools, Datasets, and Parameters

The SLAM backend was based on ORB-SLAM3, with image preprocessing implementation used OpenCV. ROS was not used in the execution pipeline. We used simple python scripts for initial prototyping and evaluation of each algorithm independently, after which the final integrated pipeline was implemented in C++ within the ORB-SLAM3 framework. All experiments were executed within a Docker container based on Ubuntu 20.04 to ensure software reproducibility. More details about build instructions and configuration files are provided in the project repository, to which, the hyperlink is attached at the end of this report.

We evaluated the DIP pipeline using custom recorded datasets which were collected in indoor corridor environments at UCSD graduate housing. Data collection took place during afternoon hours to record challenging visual conditions like strong illumination gradients and window glare. All sequences were recorded manually by a human walking through the environment while holding the camera in hand, resulting in natural camera motion and small intermittent jerks. A total of 15 to 20 runs were recorded, each approximately two minutes long. The trends that were observed for all graph plotted were consistent across all runs, and no external benchmark datasets were used. We made an attempt to evaluate our DIP pipeline on the KITTI dataset; however, limitations in adapting the dataset to the RealSense stereo configuration prevented in reliable integration.

The adaptive CLAHE module dynamically adjusted its parameters based on the image frame it received every time. Clip limits ranged from 2.0 to 6.0, and tile grid sizes varied between 4×4 and 12×12, depending on image variance and intensity range. Denoising was performed using a guided filter with window radius fixed to 16 pixels and a regularization parameter of $\epsilon = 50/255^2$. Motion blur detection used the variance of the Laplacian with a fixed threshold which ranged from 100 to 125. Motion deblurring, when activated, implemented a Wiener filter with a regularization parameter $K = 0.05$ and an IMU-guided linear motion PSF of fixed size (length = 9), where the PSF orientation was derived from integrated gyroscope yaw measurements.

Performance evaluation mainly focused on front-end and tracking-level metrics, like tracking loss frequency, keyframe insertion rate, and map point creation and culling statistics.

C. Assumptions and Constraints

The evaluation assumes that the intrinsic camera calibration parameters are accurate enough and stays stable throughout data collection. It is also assumed that the recorded environments are always static, with very low dynamic objects affecting feature tracking. All experiments assume consistent execution conditions across baseline and DIP enabled runs to allow fair comparison.

Various constraints limited the scope of the experimental evaluation. Ground truth trajectories were not available, as data was collected manually without any predefined motion paths or external tracking systems. As a result, trajectory error metrics were not computed. Also, experiments were restricted to indoor corridor environments due to time, equipment, and logistical constraints typical of a project.

These limitations shows that even though the observed improvements are consistent and reproducible within the evaluated settings, the results may not be able to generalize to all environments or extreme motion conditions. Nevertheless, the experimental setup provides a realistic and practical testing ground for analyzing the impact of adaptive front-end image processing on feature-based SLAM performance.

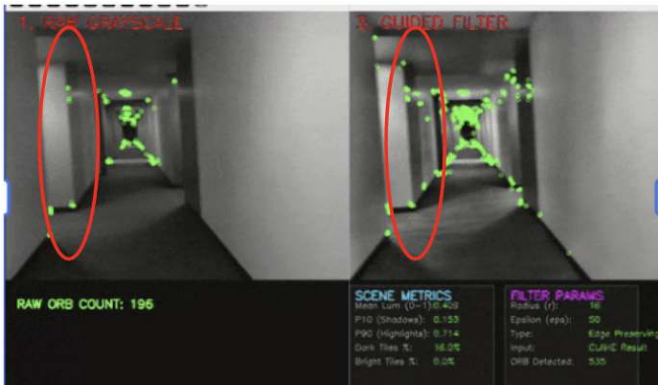


Fig. 8. Qualitative comparison illustrating the effect of the guided filter during ORB-SLAM3 operation. The left image shows baseline ORB-SLAM3 without digital image preprocessing, while the right image shows ORB-SLAM3 with the guided filter enabled as part of the adaptive DIP pipeline. The red circled regions highlight a representative corner where the guided filter enhances local contrast and edge definition, resulting in additional and more consistent feature detections. This experiment serves as a functional verification that the guided filter operates as intended within the live SLAM pipeline.

VI. Results

This section presents the quantitative and qualitative results obtained from evaluating baseline ORB-SLAM3 and the proposed adaptive DIP-enhanced pipeline. The following metrics characterize feature behavior, keyframe generation, and tracking continuity over the full sequence. Interpretation and performance comparisons are deferred to Section VII & Section VIII.

A. Map Point Creation and Culling

Figures 9 and 10 show the rolling average of map point creation and culling per keyframe for both systems. Median values are used to summarize typical behavior across the trajectory.

The baseline ORB-SLAM3 system produces a median of **154.8** map points per keyframe, while the DIP-enhanced pipeline produces a higher median of **174.0** map points per keyframe. Map point culling follows a similar trend: the baseline system exhibits a median culling rate of **258.2** points per keyframe, compared to **266.6** points per keyframe for the DIP-enhanced pipeline.

B. Keyframe Creation Frequency

Figure 11 shows the smoothed average of the spacing in the keyframes added during the run for both systems, with

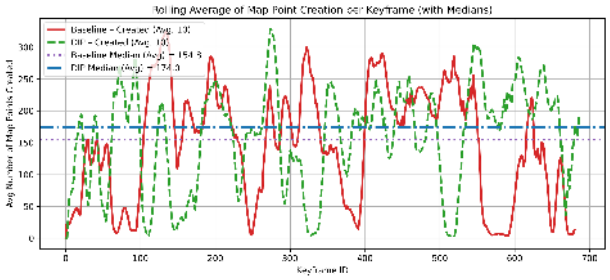


Fig. 9. Rolling average of map point creation per keyframe for baseline ORB-SLAM3 and the DIP-enhanced pipeline. Median lines indicate the typical creation rate across the trajectory.

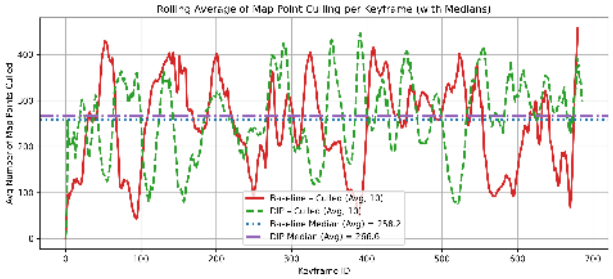


Fig. 10. Rolling average of map point culling per keyframe for baseline ORB-SLAM3 and the DIP-enhanced pipeline. Median lines summarize the typical culling behavior across the run.

median values summarizing overall behavior. The baseline system exhibits a median keyframe spacing of **7.5** frames, while the DIP-enhanced pipeline exhibits a median spacing of **6.2** frames.

C. Tracking State Timeline

Figure 12 shows the tracking state reported by ORB-SLAM3 at each frame. For the baseline system, **5,342** frames are classified as normal tracking (`OK`), while **482** frames fall into the `RECENTLY_LOST` state. The DIP-enhanced pipeline exhibits **5,501** frames in the `OK` state and **369** frames in the `RECENTLY_LOST` state. Neither system enters a persistent `LOST` tracking state during the evaluated sequence.

VII. Comparison with Baseline ORB-SLAM3

In this section we interpret the results presented in Section V by comparing the behavior of the DIP-enhanced pipeline against baseline ORB-SLAM3. The comparison focuses on the feature quality, tracking stability and overall system robustness.

A. Feature Quality and Map Point Behavior

The DIP-enhanced pipeline created a higher median number of map points per keyframe as compared to the baseline system (174.0 vs. 154.8). This increase indicates that the improved image quality by the adaptive DIP front-end enables the detection and triangulation of more number of usable visual

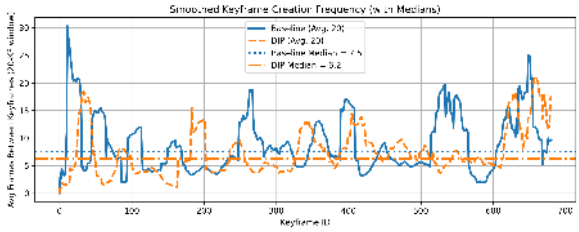


Fig. 11. Smoothed keyframe frequency shows the frequency at which the keyframes are being generated for baseline ORB-SLAM3 and the DIP-enhanced pipeline. Median lines summarize overall behavior.

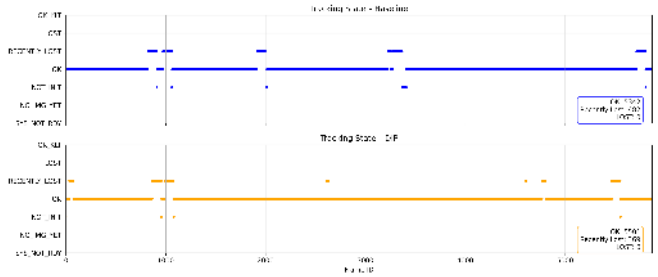


Fig. 12. Tracking state timeline for baseline ORB-SLAM3 (top) and the DIP-enhanced pipeline (bottom). Each marker indicates the reported tracking state at a given frame.

features, particularly in frames affected by noise, motion blur and challenging illumination.

At the same time, the DIP pipeline exhibits a slightly higher median map point culling rate (266.6 vs. 258.2). Although the increase in culling is modest, it suggests more active map maintenance, where newly created landmarks are evaluated and removed if they fail to meet ORB-SLAM3's geometric consistency. Importantly, the increase in culling does not offset the higher creation rate, indicating that the system is maintaining a larger pool of good landmarks overall.

Taken together, these results suggest that the adaptive DIP front-end improves feature detectability and at the same time preserves the selectivity of the SLAM back-end. Rather than indiscriminately increasing map density, the pipeline enables the initialization of additional candidate landmarks while still enforcing strict consistency checks. This behavior shows an improvement in feature reliability, where enhanced image quality leads to more frequent initialization of meaningful landmarks without allowing unstable features to persist in the map.

B. Keyframe Stability

The median spacing between consecutive keyframes differs modestly between the two systems, with the DIP-enhanced pipeline exhibiting a slightly lower median spacing (6.2 frames) compared to the baseline system (7.5 frames). This corresponds to a small increase in keyframe insertion frequency.

When interpreted alongside the map point statistics, this behavior suggests that improved feature quality results in more

reliable and consistent keyframe creation rather than reactive keyframe insertion driven by the tracking degradation. The higher availability of stable and repeatable map points allows ORB-SLAM3 to anchor the map using well-conditioned keyframes that are inserted under normal tracking conditions, rather than as a response to feature loss or impending relocalization threat.

C. Tracking Continuity

Tracking state timelines highlight the robustness of the DIP-enhanced pipeline. The DIP system remains in the normal, "OK", tracking state for 5,501 frames, compared to 5,342 frames for the baseline, while reducing the number of frames spent in degraded tracking (`RECENTLY_LOST`) from 482 to 369.

This reduction in degraded tracking time indicates improved frame-to-frame feature consistency and greater robustness for challenging visual conditions such as motion blur and lighting transitions. Although neither system experiences complete tracking failure, the DIP-enhanced pipeline demonstrates more consistent tracking behavior during the testing.

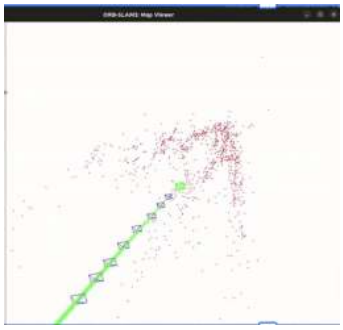


Fig. 13. Three-dimensional map visualization produced by ORB-SLAM3 during testing with the proposed adaptive DIP front-end. Red points represent triangulated 3D map landmarks reconstructed from tracked visual features. Blue rectangular markers represent keyframes stored in the map, which serve as reference frames for local mapping, loop closure, and relocalization. The green continuous trajectory indicates the estimated camera pose over time. This visualization illustrates stable map growth and consistent pose estimation throughout the test.

VIII. Discussion

The experimental results demonstrate that integrating an adaptive DIP front-end into ORB-SLAM3 yields consistent, though moderate, improvements in feature quality and tracking robustness. Even though the core SLAM architecture remains unchanged, the observed improvements highlight the importance of front-end image quality in feature-based localization and mapping systems.

A. Interpretation of Results

The map point creation and culling results indicate that the DIP enhanced pipeline improves feature detectability while preserving strict back-end consistency checks. The DIP pipeline produces a median number of 174.0 map points per keyframe compared to the baseline of 154.8, suggesting

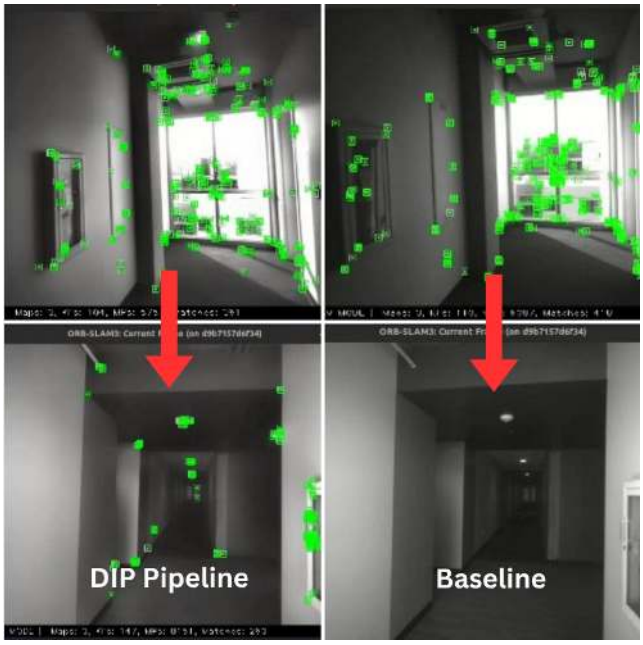


Fig. 14. Qualitative comparison of tracking behavior under strong window glare. The top row shows frames captured immediately before a corridor turn when facing a bright window, while the bottom row shows the corresponding frames after the turn. The left column corresponds to ORB-SLAM3 with the adaptive DIP front-end, and the right column shows the baseline system. In the baseline case, severe glare causes a loss of detectable features, leading to tracking failure and subsequent drift until loop closure occurs. In contrast, the DIP-enhanced pipeline preserves sufficient feature structure through contrast enhancement and denoising, enabling continuous tracking and preventing drift throughout the sequence.

that CLAHE, denoising, and deblurring enable the system to initialize additional usable feature candidates, particularly in frames affected by noise, motion blur, or dynamic lighting.

At the same time, the DIP-enhanced pipeline exhibits a median culling of 266.6 map points, while the base line culled a median of 258.2. Although this increase is modest, it demonstrates more active map maintenance, where newly initialized landmarks are evaluated and removed if they fail to satisfy ORB-SLAM3's geometric and visibility constraints. Importantly, the slightly higher culling rate does not negate the higher creation rate, which implies that the system maintains a larger pool of total viable landmarks overall. This behavior reflects improved feature reliability rather than an indiscriminate increase map point density.

The frequency of keyframe creation further supports this interpretation when considered in conjunction with improved map point statistics. Although the observed differences in median keyframe frequency are relatively small, they indicate that the DIP pipeline does not fundamentally alter ORB-SLAM3's keyframe defining strategy. Instead, the availability of more reliable map points enables keyframes to be inserted under stable tracking conditions rather than creating keyframes for relocalization. This suggests reduced dependence on keyframes as recovery mechanisms, even though the overall insertion rate remains similar.

The tracking state timeline provides confirmation of these trends. The DIP enhanced system spends a greater proportion of the sequence in the normal tracking state and reduces the total time spent in degraded tracking states. Although neither system experiences consistent tracking loss, the reduced tracking lost state indicates improved robustness when operating under visually degraded conditions.

B. Trade-offs and Limitations

The primary limitation of the proposed adaptive DIP front-end lies in its dependence on parameter tuning specific to certain environmental conditions. The parameters used in this study, which included contrast enhancement thresholds, guided filter window sizes, and blur detection criteria, were empirically tuned for indoor corridor environments with moderate lighting transitions. While these settings produced consistent improvements under the evaluated conditions, they may potentially falter under significantly different conditions, such as outdoor scenes with extreme illumination changes or low-texture industrial settings.

As a result, utilizing the pipeline in varying environments would likely require re-tuning of preprocessing parameters. Without appropriate tuning, overly aggressive contrast enhancement or denoising could suppress weak, but useful features. While overly conservative settings may fail to mitigate visual degradation. This highlights an inherent trade-off in classical image processing approaches, where robustness depends on environment-aware parameter selection rather than learned generalization.

Additionally, the improvements observed in this study are incremental rather than transformative. The DIP pipeline does not completely eliminate tracking degradation, neither does it fundamentally alter ORB-SLAM3's keyframe insertion strategy. Instead, it provides consistent gains that accumulate across the pipeline, improving feature reliability and tracking continuity without changing the underlying SLAM architecture. This outcome demonstrates both the importance of front-end image quality in feature-based SLAM and the inherent limits of preprocessing in addressing all sources of tracking instability.

C. Practical Implications for Real-World Deployment

From a real-world deployment perspective, the results demonstrate that adaptive image preprocessing can extend the operational robustness of feature-based SLAM systems in environments where lighting, motion dynamics, and sensor noise cannot be controlled. The observed improvements in feature reliability and tracking continuity are particularly important in scenarios where intermittent tracking degradation can be more disruptive than brief moments of tracking loss.

The modular nature of the proposed DIP front-end allows it to be integrated as a plug-in enhancement to existing SLAM systems without the intrusive modification of the back-end. This makes the approach suitable for real-time robotic applications that require interpretability, efficiency, and ease of implementation. Overall, the findings suggest that classical, adaptive digital image processing remains practical and effective for improving the robustness of modern visual SLAM systems when applied in a targeted and systematic manner.

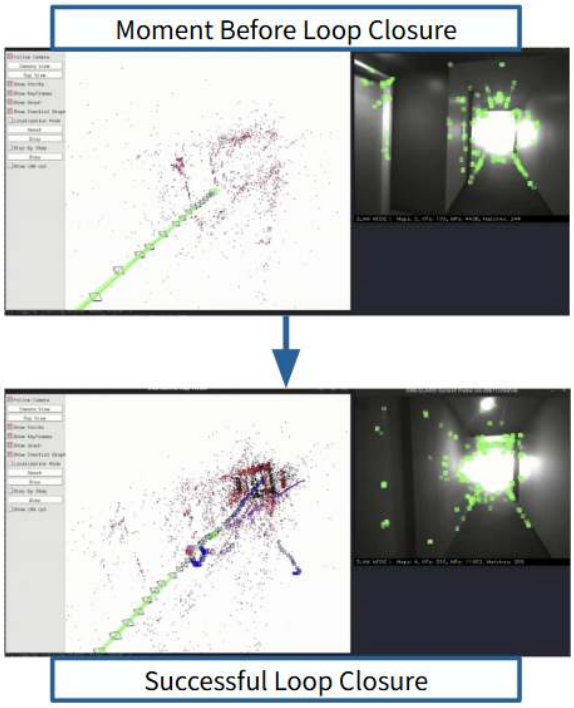


Fig. 15. Qualitative visualization of loop closure behavior with the adaptive DIP front-end enabled. Both panels show ORB-SLAM3 operating with the proposed DIP pipeline. The top panel corresponds to the moment immediately before loop closure, where drift is present despite continuous tracking. The bottom panel shows the state after loop closure.

IX. Future Work

While the proposed adaptive digital image processing pipeline demonstrates slightly improved robustness of ORB-SLAM, several important directions remain for future investigation.

First, there are fundamental limitations to classical digital image processing techniques when dealing with severe motion blur. In cases of rapid camera motion, blur can be highly spatially varying and non-linear, making it difficult for classical deconvolution methods to fully restore geometrically consistent image content. Even with estimated point spread functions, aggressive deblurring can introduce artifacts that degrade feature repeatability across frames. Addressing these cases may require alternative modeling assumptions or more advanced blur-aware processing strategies beyond classical deconvolution.

Second, the performance of IMU-guided motion deblurring was constrained by uncertainties in the overall sensing and estimation pipeline. Although inertial measurements were calibrated, the system continued to exhibit drift and inconsistent deblurring results, suggesting that calibration alone was insufficient to resolve the issue. These limitations may stem from

a combination of factors, including sensor noise, imperfect camera IMU synchronization, rolling shutter effects, or reduced SLAM stability in visually degraded environments. As a result, inaccuracies in motion estimation directly impacted point spread function estimation and limited the effectiveness of deblurring. Addressing this limitation may require higher-quality or more tightly synchronized sensing hardware, or alternatively, algorithmic approaches that reduce reliance on inertial measurements that are more robust to uncertainty in motion estimation.

Third, the pipeline should be evaluated in more diverse and challenging real-world environments, such as industrial facilities, dense public spaces, or underground mines. These environments often contain low illumination, high dynamic range lighting, repetitive textures, and airborne particulates, all of which pose significant challenges for visual SLAM systems. Evaluating performance in such settings would provide a more comprehensive understanding of the pipeline's practical robustness and limitations.

Fourth, the observed improvements suggest diminishing returns from front-end-only preprocessing. While enhancing image quality improves feature detection and tracking stability, overall SLAM performance is jointly determined by a culmination of the front-end, tracking, and back-end optimization components. Achieving larger system-level gains may require coordinated improvements that extend beyond image preprocessing which may include modifications to back-end optimization or state estimation strategies.

Finally, future evaluation would benefit from trajectory-level accuracy metrics such as Absolute Trajectory Error or Relative Pose Error. In this work, such metrics could not be computed due to the absence of a ground truth, as data was collected manually without a predefined trajectory. Generating large-scale datasets with accurate ground truth would require extensive infrastructure and time beyond the scope of this project, but remains an important direction for future validation.

X. Conclusion

This project investigated the impact of an adaptive digital image processing front-end on the robustness of ORB-SLAM under visually challenging conditions. By selectively applying contrast enhancement, edge-preserving denoising, and motion deblurring, the proposed pipeline improves the quality of visual features delivered to the SLAM front-end without modifying the core ORB-SLAM architecture.

Experimental results demonstrate reduced tracking loss, lower keyframe frequency rates, and improved map point retention, which indicated more stable and consistent front-end behavior. These improvements, while incremental, confirm that front-end image quality plays a critical role in the performance of feature-based SLAM systems, particularly in environments with degraded visual conditions.

Overall, this work shows that classical, digital image processing techniques remain effective tools for improving SLAM robustness when applied adaptively and with system-level awareness. Although front-end preprocessing alone cannot address all failure modes, the proposed pipeline provides a lightweight foundation for future research into more comprehensive, SLAM-aware perception systems.

Supplementary Material

The source code for this project, including the adaptive digital image processing pipeline with ORB-SLAM3 integration, and experimental configuration files, is publicly available on Github at:

https://github.com/daatsi-aeres/DIP_ORB_SLAM3_ .

Declarations

Conflict of Interest

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' Contributions

Siddharth Rama Sushil, Shahid Mohammad Mulla, and Guillermo Lievano contributed equally to this project. All authors were involved in the design, implementation, experimentation, analysis, and writing of the manuscript.

TABLE I
Effort distribution among authors

Author	Contribution (%)
Siddharth Rama Sushil	33.33%
Shahid Mohammad Mulla	33.33%
Guillermo Lievano	33.33%

References

[1] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, and J. D. T. Sastre, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[3] P. Savinykh, B. M. L. K. K. P. Herath, and E. S. L. P. E. Jayaweera, "DarkSLAM: GAN-based low-light enhancement for vSLAM," in *Proc. 2022 International Conference on Mechatronics, Robotics and Automation (ICMRA)*, 2022, pp. 10–15.

[4] C. Yu, Z. Liu, X. J. Liu, F. Xie, Y. Yang, Q. Wei, and Q. Fei, "DS-SLAM: A Semantic Visual SLAM for Dynamic Environments," in *Proc. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 1618–1624.

[5] D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperPoint: Self-Supervised Interest Point Detection and Description," in *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018, pp. 336–33612.

[6] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. Academic Press, 1994, pp. 474–485.

[7] "Histogram Equalization," OpenCV Documentation. Accessed: Nov. 6, 2025. [Online]. Available: https://docs.opencv.org/4.x/d5/daf/tutorial_py_histogram_equalization.html

[8] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[9] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Sixth International Conference on Computer Vision*, 1998, pp. 839–846.

[10] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral Filtering: Theory and Applications," *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, pp. 1–73, 2009.
[11] J. Jia, "Single image motion deblurring using transparency," in *Proc. 2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
[12] L. Liu, Y. Wang, and Z. Zhang, "IMU-aided adaptive mesh-grid based video motion deblurring," *Scientific Reports*, vol. 14, 2024.
[13] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski, "Image deblurring using inertial measurement sensors," *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 1–9, 2010.
[14] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.
[15] K. He, J. Sun, and X. Tang, "Guided Image Filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 35, no. 6, pp. 1397–1409, 2013.