



Wavelet-based image preprocessing for enhancing trajectory estimation in monocular ORB-SLAM3

Adrian Kapski^{*}, Paweł Słowak, Szymon Płatek, Piotr Kaniewski

Institute of Radioelectronics, Faculty of Electronics, Military University of Technology
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland

Article info:

Article history:

Received 23 Jun. 2026
Received in revised form 17 Jul. 2026
Accepted 03 Aug. 2026
Available on-line 09 Sep. 2026

Keywords:

monocular SLAM;
ORB-SLAM3;
wavelet preprocessing;
denoising;
trajectory error.

Abstract

This paper examines whether classical wavelet-based image preprocessing can improve the accuracy of camera-based simultaneous localisation and mapping (SLAM). In such systems, a moving camera estimates its own trajectory by repeatedly recognising visual structures in consecutive images. The quality of these images therefore has a direct influence on the stability of localisation, especially when the scene contains noise, weak texture, uneven illumination, or rapidly changing visual details. The applied approach involves preprocessing the input images before they are passed to the localisation algorithm, while keeping the remaining conditions unchanged. Wavelet preprocessing is used because it separates an image into components representing coarse brightness information and finer structural details such as edges, junctions, and textures. This makes it possible to reduce unstable fine-scale fluctuations while preserving image structures that are useful for visual localisation. Five interpretable wavelet-based preprocessing variants are compared with the original monocular image stream using repeated evaluations on standard camera-trajectory sequences with reference motion data. The results show that wavelet preprocessing can improve trajectory estimation when it strengthens stable visual information without distorting the scene structure. The best variant reduces the mean absolute trajectory error by 26.1% compared with the unprocessed monocular baseline. At the same time, the results indicate that the effectiveness of preprocessing depends on how well the transformation preserves information relevant to tracking under demanding motion and scene conditions. The study therefore presents wavelet preprocessing as a low-cost, interpretable image-side support for monocular visual SLAM, together with its practical limitations.

1. Introduction

Reliable estimation of a camera trajectory from a single image stream is an important requirement for modern unmanned platforms, including unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), and unmanned underwater vehicles (UUVs). Visual simultaneous localisation and mapping, commonly referred to as visual SLAM, estimates the camera motion while simultaneously reconstructing a map of the observed environment. In monocular visual SLAM, only one camera is available; therefore, both camera motion and scene structure must be inferred from a single image stream. As a result, localisation accuracy strongly depends on the quality, stability, and repeatability of visual information extracted from consecutive frames [1–5].

Among feature-based visual SLAM systems, ORB-SLAM3 is widely used as a reference implementation because of its accuracy, open-source design, and suitability for controlled experimentation [6]. ORB denotes Oriented FAST and Rotated BRIEF. FAST, or Features from Accelerated Segment Test, is a corner detector based on local pixel-intensity comparisons. BRIEF, or Binary Robust Independent Elementary Features, is a binary descriptor that represents a small image patch using intensity comparisons. In ORB-SLAM3, camera motion is therefore estimated from image points that must be detected, described, matched, and geometrically verified across consecutive frames [7].

This makes the input image a measurement signal for the SLAM algorithm. Pixel-level noise, weak texture, uneven illumination, local contrast changes, and motion blur can affect both FAST keypoint detection and BRIEF-like

^{*}corresponding author at: adrian.kapski@wat.edu.pl

descriptor responses, even when the same physical point is observed. Consequently, the number and stability of features, the quality of feature matching, and the accuracy of geometric constraints used for trajectory estimation may change [8–10].

This paper investigates whether wavelet-domain image preprocessing can improve monocular ORB-SLAM3. Wavelet processing is well suited to this task because it decomposes an image into components associated with different spatial scales. The approximation component represents coarse brightness variation, whereas detail components contain information about edges, corners, and textures [11, 12]. Classical wavelet-denoising methods, such as soft thresholding, BayesShrink, SureShrink, and local adaptive shrinkage, provide established mechanisms for suppressing noise-like coefficients while preserving meaningful image structures. These mechanisms are relevant to ORB-based SLAM because stable localisation depends on repeatable local intensity relationships used during feature detection, description, and matching.

The study compares five wavelet-based preprocessing variants, denoted W1–W5, covering suppression of unstable fine-scale fluctuations, reinforcement of image structures, adaptive subband denoising, and modification of the low-frequency brightness component. All variants are evaluated against the original monocular image stream using the same camera calibration, timestamps, vocabulary, feature-extractor settings, and localisation algorithm. The evaluation includes trajectory error, result stability, tracking continuity, runtime, the number of keyframes, and the number of map points. The main contribution of this work is a controlled assessment of classical wavelet mechanisms as image-side support for an unchanged ORB-SLAM3 algorithm.

2. Related work

2.1. Visual SLAM and image-dependent localisation

Visual SLAM systems differ in how they use image information for motion estimation. Feature-based methods, such as ORB-SLAM2 and ORB-SLAM3, detect and match keypoints between consecutive frames and use the resulting correspondences for geometric estimation and graph optimisation [6, 7]. Direct and semi-direct methods, such as large-scale direct (LSD)-SLAM and direct sparse odometry (DSO), rely on photometric consistency between images [11, 12], whereas visual-inertial systems combine camera and inertial measurement unit (IMU) measurements to improve motion estimation under fast motion or temporary visual degradation [13].

In ORB-based algorithms, FAST detection and BRIEF description rely on local intensity comparisons. Consequently, illumination changes, noise, blur, and weak texture may affect feature detection, matching quality, and trajectory-estimation accuracy [8–10]. This motivates image conditioning before feature extraction to improve the geometric information available to the SLAM algorithm.

2.2. Image preprocessing in SLAM

Image preprocessing has been used to improve SLAM performance under visually degraded conditions. Active front end (AFE)-ORB-SLAM combines image enhancement with adaptive FAST thresholding [14], whereas Dark-SLAM targets low-light environments [15]. Other methods improve brightness, contrast, and structural visibility in underwater or weak-texture scenes [16, 17]. However, these approaches are mainly designed for specific environmental conditions or modify feature-extraction mechanisms, rather than isolating the influence of different classical wavelet operations within an unchanged ORB-SLAM3 pipeline.

Classical image-processing operations, including sharpening and edge enhancement, can also be applied before feature extraction in visual SLAM. Sharpening may strengthen local intensity gradients and improve the visibility of corners and structural boundaries. However, excessive sharpening can amplify noise and introduce artificial high-frequency structures, while edge-only representations may remove texture information required for reliable feature description and matching. Wavelet preprocessing provides greater scale selectivity, allowing selected detail components to be modified while retaining the complete image structure [8–10].

2.3. Wavelet-based processing in visual localisation and SLAM

Wavelet representations have previously been used to construct compact signatures for topological localisation [18], detect blurred frames before pose estimation [19], and design feature descriptors [20]. In these studies, wavelets support place recognition, frame selection, or descriptor construction. In contrast, the present work applies wavelet processing directly to every input frame before ORB feature extraction and evaluates its influence on the complete monocular ORB-SLAM3 pipeline.

2.4. Research gap and contribution

The reviewed literature lacks a controlled, end-to-end comparison of interpretable wavelet preprocessing mechanisms applied directly to the input images of monocular ORB-SLAM3. To address this gap, five variants, denoted W1–W5, are selected to represent fine-scale denoising, structural-detail reinforcement, adaptive subband thresholding, and approximation-component modification [21–25]. Unlike previous methods that modify illumination handling, frame selection, feature descriptors, or elements of the SLAM pipeline, the proposed study modifies only the input images while preserving the original ORB-SLAM3 algorithm and its configuration. The contribution of this study is therefore a controlled end-to-end comparison that isolates the influence of five interpretable wavelet mechanisms on trajectory accuracy, repeatability, tracking continuity, runtime, and map construction, without modifying the ORB-SLAM3 algorithm or its configuration.

3. Materials and methods

3.1. Base system, dataset, and experimental design

The baseline system was monocular ORB-SLAM3 [6], used in its publicly released form without any code-level modification. The camera model, EuRoC calibration file, ORB vocabulary, feature-extractor settings, timestamps, and ORB-SLAM3 itself were kept unchanged. No feature threshold, local mapping parameter, vocabulary file, camera parameter, or ORB-SLAM3 optimisation setting was retuned for any preprocessing variant. All executions were performed on the same computational platform and software stack. For every tested variant, the input sequence retained the same frame geometry and temporal order as the original benchmark. The experiment therefore isolates photometric and scale-domain image preprocessing from changes in the SLAM estimator.

The benchmark was the complete EuRoC MAV dataset [26]. EuRoC provides synchronised MAV image sequences and ground-truth trajectories for standard SLAM evaluation; in this paper only the monocular camera stream was passed to ORB-SLAM3. The dataset is divided into two groups. Machine Hall sequences MH01–MH05 were recorded in a larger industrial hall, where the camera observes machinery, structural elements, and longer-range scene geometry. The first Machine Hall sequences are easier, while the later ones contain more demanding motion and visual conditions. Vicin Room sequences V101–V103 and V201–V203 were recorded in smaller motion-capture rooms, where the camera observes closer objects, shorter scene depth, stronger viewpoint changes, and less room for stable parallax than in the hall. The Vicin sequences therefore differ from Machine Hall not only in appearance, but also in the geometry available to a monocular estimator. In the present results, V203 is the most difficult Vicin Room case: it combines aggressive motion, rapid viewpoint changes, and a high risk of motion blur and tracking loss. These differences are important because a preprocessing method that works well for moderate image noise or contrast variation may fail when the dominant limitation is camera motion and geometry.

The evaluated set contains the unprocessed monocular baseline and five wavelet-preprocessed image streams. Each method was run 10 times on each of the 11 EuRoC sequences:

$$N_{\text{runs}} = 6 \times 11 \times 10 = 660. \quad (1)$$

The repeated-run protocol is necessary because monocular ORB-SLAM3 can produce slightly different trajectories across nominally identical executions. The paper therefore reports sequence-level means and dispersion over repeated runs, not best-run examples.

3.2. Wavelet preprocessing variants

All five preprocessing variants use the same outer pipeline: each grayscale frame is converted to floating point, decomposed by a two-dimensional discrete wavelet transform, modified in selected coefficient bands, reconstructed by the inverse transform, cropped back to the original image size if padding was required, and clipped to the 8-bit range. The Daubechies-4 basis was selected because compactly supported wavelets provide localised multiresolution image representation, which allows fine-scale fluctuations to be separated from broader edges, corners, junctions, and texture structures [21]. The decomposition depths $L = 2$ and $L = 3$ were chosen to cover fine and medium-scale structures that are relevant to FAST keypoint detection and BRIEF-like descriptor construction, both of which rely on local intensity comparisons. The variants therefore differ only in decomposition depth and coefficient-domain operation, and they were selected to represent distinct mechanisms known from wavelet denoising and multiresolution analysis: soft thresholding, BayesShrink, SureShrink, detail reinforcement, and approximation-band contrast modification [22–25]. Table 1 summarises them with the compact symbols W1–W5 used throughout the rest of the paper. Thus, W1–W5 were not selected as arbitrary parameter combinations; each variant represents a distinct hypothesis concerning which spatial-scale information is most beneficial for ORB feature extraction and trajectory estimation.

The physical interpretation of the wavelet bands is central to the experiment. The approximation band represents slow

Table 1

Wavelet preprocessing variants. All variants use the Daubechies-4 (db4) basis; only the decomposition depth L and the coefficient-domain operation differ. Gain $g=1.3$.

Variant	Coefficient-domain operation	Intended mechanism
W1 ($L=3$)	BayesShrink on level-1 details; gain g on level-2 and level-3 details; approximation unchanged.	Suppresses pixel-scale fluctuations and reinforces medium-scale edges and junctions; designed for repeatable feature behaviour.
W2 ($L=2$)	Universal soft thresholding followed by gain g on level-1 and level-2 details; approximation unchanged.	Couples denoising with controlled detail amplification; tests whether joint denoising and controlled detail amplification improves feature stability and trajectory estimation.
W3 ($L=3$)	Independent BayesShrink on all level-1 to level-3 details; no gain.	Conservative scale-aware denoising that attenuates noise-like detail without explicit detail reinforcement.
W4 ($L=3$)	CLAHE applied only to the level-3 approximation; all detail bands unchanged.	Modifies low-frequency brightness only; isolates the effect of regional contrast from detail conditioning.
W5 ($L=2$)	SureShrink (SURE-based soft thresholding) on level-1 and level-2 details; no gain.	Risk-based detail denoising that adapts the threshold to each band without a fixed global value.

illumination and brightness variation. The finest detail bands contain the smallest image fluctuations, including noise-like pixel-scale changes. Coarser detail bands contain broader edges, junctions, and medium-scale texture structures that are more likely to support repeatable FAST corners and ORB descriptors. The five variants therefore compare different ways of redistributing image information across scales before ORB-SLAM3 performs feature extraction.

3.3. Frame-level visual comparison

Figure 1 provides an image-domain interpretation of the five coefficient-domain definitions before the mathematical model is introduced. The overview row marks two regions of interest, and the lower rows show the same regions after W1–W5. This figure is not a performance metric; it explains what type of visual evidence is delivered to the unchanged ORB feature extractor. W1 and W2 preserve stronger structural detail after denoising, W3 and W5 behave more conservatively, and W4 changes local brightness distribution more visibly because it acts on the approximation band. This is why the trajectory analysis later distinguishes useful scale-selective detail conditioning from visually stronger but less reliable image changes.

3.4. Mathematical model of the preprocessing

Let $I(x, y)$ be an 8-bit grayscale image, where x and y are pixel coordinates and I denotes the image-intensity function. Following the multiresolution representation of Mallat [21], the discrete wavelet transform separates this image into one low-frequency image and several detail images:

$$I \xrightarrow{\text{DWT}_{\text{db4},L}} \{A_L, (H_l, V_l, D_l)_{l=1}^L\}. \quad (2)$$

In (2), $\text{DWT}_{\text{db4},L}$ denotes a two-dimensional discrete wavelet transform using the Daubechies-4 basis and decomposition depth L . The symbol A_L is the approximation image at level L and contains slowly varying brightness and coarse scene content. For every level $l \in \{1, \dots, L\}$, the three images H_l , V_l , and D_l contain horizontal, vertical, and diagonal detail, respectively. Level $l = 1$ contains the finest image detail, while larger values of l describe broader structures. This notation is used only for wavelet subbands; D_l means diagonal detail, not a derivative.

Each method modifies selected coefficient images and reconstructs the image:

$$\tilde{I}_m = \text{clip}_{[0,255]} (\text{IDWT}(\tilde{A}_L, \tilde{H}_l, \tilde{V}_l, \tilde{D}_l)), \quad (3)$$

where m denotes the preprocessing variant, $\tilde{I}_m$ is the reconstructed image delivered to ORB-SLAM3 for variant m , IDWT is the inverse wavelet transform, and the tilde marks a modified coefficient image. The operator $\text{clip}_{[0,255]}(\cdot)$ limits reconstructed intensities to the valid 8-bit range after inverse transformation.

For compact notation, let B_l denote any one of the three detail images at level l , that is $B_l \in \{H_l, V_l, D_l\}$. The thresholding rules used below are based on the standard additive coefficient model:

$$z_{l,k} = s_{l,k} + \eta_{l,k}, \quad \eta_{l,k} \sim \mathcal{N}(0, \sigma_n^2), \quad (4)$$

where $z_{l,k}$ is the observed detail coefficient, $s_{l,k}$ is the unknown structural component, and $\eta_{l,k}$ is the noise. This is a working model used to derive thresholding rules; it is not an assertion that all EuRoC image coefficients are independent Gaussian measurements. The model is the usual starting point for wavelet shrinkage in image denoising [22–24]. Denoising then uses soft thresholding:

$$\text{soft}(z, T) = \begin{cases} 0, & |z| \leq T, \\ \text{sign}(z)(|z| - T), & |z| > T. \end{cases} \quad (5)$$

Here z is one wavelet coefficient, T is the non-negative threshold, $|z|$ is coefficient magnitude, and $\text{sign}(z)$ preserves the coefficient sign after shrinkage. Coefficients with small magnitude are removed, while stronger coefficients are reduced but preserved. This follows the soft-shrinkage rule of Donoho and Johnstone [22, 23]; in the present application it suppresses noise-like detail without completely erasing edges.

The noise level is estimated from the finest diagonal detail image:

$$\hat{\sigma}_n = \frac{\text{median}(|D_1|)}{0.6745}. \quad (6)$$

In (6), $\hat{\sigma}_n$ is the estimated noise standard deviation and $\text{median}(|D_1|)$ is the median absolute value of coefficients in the finest diagonal subband. The denominator 0.6745 is the standard median absolute deviation conversion for zero-mean Gaussian noise and is commonly used to estimate the noise level from the finest diagonal subband [22, 23]. The three thresholding rules used in this study are

$$T_U(l) = \hat{\sigma}_n \sqrt{2 \ln M_l}, \quad (7)$$

$$T_B(l) = \frac{\hat{\sigma}_n^2}{\hat{\sigma}_{S,l} + \varepsilon}, \quad \hat{\sigma}_{S,l} = \sqrt{\max(\hat{\sigma}_{B,l}^2 - \hat{\sigma}_n^2, 0)}, \quad (8)$$

$$T_S(l) = \arg \min_T \hat{R}_l(T). \quad (9)$$

In (7), (8), (9), M_l is the number of coefficients in the processed detail image, $\hat{\sigma}_{B,l}$ is the measured standard deviation of that detail image, $\hat{\sigma}_{S,l}$ is the estimated standard deviation of the clean structural detail, ε prevents division by zero, and $\hat{R}_l(T)$ is the Stein unbiased risk estimate for soft thresholding. Equation (7) is the universal threshold used by W2 [22, 23]. Equation (8) is the BayesShrink threshold used by W1 and W3 [24]. Equation (9) is the SureShrink choice used by W5 [23]. Thus, every threshold in the experiment is tied to a standard wavelet-denoising rule rather than fitted to the EuRoC results.

With gain $g = 1.3$, the five tested variants are:

$$\begin{aligned} \text{W1 : } & B'_1 = \text{soft}(B_1, T_B(1)), \\ & B'_2 = gB_2, \quad B'_3 = gB_3, \quad A'_3 = A_3, \\ \text{W2 : } & B'_l = g \text{soft}(B_l, T_U(l)), \quad l \in \{1, 2\}, \\ & A'_2 = A_2, \\ \text{W3 : } & B'_l = \text{soft}(B_l, T_B(l)), \quad l \in \{1, 2, 3\}, \\ & A'_3 = A_3, \\ \text{W4 : } & A'_3 = \text{CLAHE}(A_3), \\ & B'_l = B_l, \quad l \in \{1, 2, 3\}, \\ \text{W5 : } & B'_l = \text{soft}(B_l, T_S(l)), \quad l \in \{1, 2\}, \\ & A'_2 = A_2. \end{aligned} \quad (10)$$

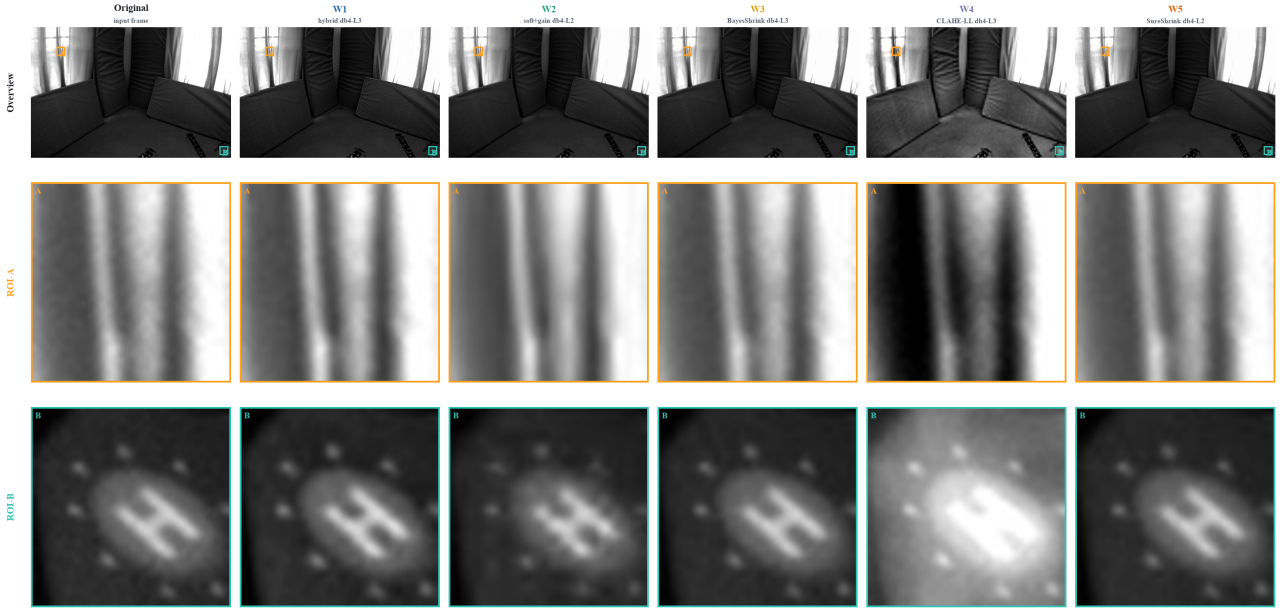


Fig. 1. Frame-level comparison on one shared Vicon Room input frame. Coloured boxes in the overview row mark the ROIs enlarged below. Method symbols W1–W5 are defined in Table 1.

Primes in (10) denote the modified coefficient images used during reconstruction. For every detail level, the operation is applied independently to the horizontal, vertical, and diagonal detail images. The definitions show the intended mechanisms. W1 and W2 combine suppression of weak coefficients with reinforcement of remaining detail. W3 and W5 perform denoising without gain. W4 changes only coarse brightness structure and leaves detail coefficients unchanged.

3.5. Connection to FAST and ORB descriptors

The relevance to ORB-SLAM3 follows from the ORB feature-extraction stage. FAST accepts a corner candidate at pixel p when enough pixels q_k on a circle around p are sufficiently brighter or darker than p [8]. The index k enumerates pixels on the FAST test circle. The basic contrast quantity is

$$c_I(p, q_k) = |I(q_k) - I(p)|. \quad (11)$$

In (11), $c_I(p, q_k)$ is the absolute intensity contrast between the candidate pixel p and circle pixel q_k in the original image I . The candidate is supported when this contrast exceeds the FAST threshold for a sufficient contiguous set of circle pixels. After preprocessing, ORB-SLAM3 sees $\tilde{I}_m$ instead of I , so the same test uses

$$c_{\tilde{I}_m}(p, q_k) = |\tilde{I}_m(q_k) - \tilde{I}_m(p)|. \quad (12)$$

This gives a direct interpretation of the wavelet operations. Denoising can remove random threshold crossings caused by fine-scale fluctuations. Moderate detail gain can increase the contrast margin of stable edges and junctions. Excessive modification can have the opposite effect by creating unstable corners or changing their relative strength. ORB descriptors are also intensity-comparison descriptors built from

BRIEF-like binary tests [9, 10]. A single binary comparison can be written as

$$b_j(\tilde{I}_m) = \mathbb{I}\{\tilde{I}_m(u_j) < \tilde{I}_m(v_j)\}, \quad (13)$$

where b_j is the j th descriptor bit, u_j and v_j are two sampling locations in a local patch, and $\mathbb{I}\{\cdot\}$ is the indicator function, equal to 1 when the condition inside braces is true and 0 otherwise. The bit remains unchanged as long as the intensity ordering between u_j and v_j does not change. A useful preprocessing operation should therefore suppress intensity fluctuations that can invert such local comparisons, while preserving or moderately strengthening image structures that are repeatable across neighbouring frames and viewpoints. In this sense, the FAST and ORB equations provide a mechanistic link between scale-selective image conditioning and the trajectory-level behaviour observed later in the experiments.

3.6. Evaluation protocol

The evaluation follows the standard trajectory-comparison logic used for EuRoC-based SLAM studies [6, 26]. Reference-based validation has also been applied to other positioning systems, including experiments in which controlled reference motion was reproduced using a multi-degree-of-freedom industrial robot [27]. Because monocular trajectories have an arbitrary scale, every estimated trajectory is aligned to ground truth by a similarity transformation, Sim(3), before error computation. The Sim(3) transformation accounts for rotation, translation, and scale. For aligned positions $\mathbf{p}_i^{gt}$ and $\mathbf{p}_i^{est}$, where gt denotes ground truth, est denotes the estimate, and i is the trajectory-sample index, the absolute trajectory error at sample i is

$$e_i^{ATE} = \|\mathbf{p}_i^{gt} - \mathbf{p}_i^{est}\|_2, \quad (14)$$

where $\|\cdot\|_2$ is the Euclidean norm. The reported ATE RMSE is

$$ATE_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i^{ATE})^2}. \quad (15)$$

In (15), n is the number of associated trajectory samples used after timestamp matching. Coverage is

$$\text{coverage} = \frac{N_{\text{assoc}}}{N_{\text{cam}}}, \quad (16)$$

where N_{assoc} is the number of estimated poses associated with ground truth and N_{cam} is the number of camera timestamps in the evaluated sequence.

For method m and sequence s , trajectory improvement is reported relative to the 10-run baseline mean on the same sequence:

$$\text{Imp}(m, s) = 100 \frac{\mu_{\text{base},s} - \mu_{m,s}}{\mu_{\text{base},s}}, \quad (17)$$

where $\mu_{\text{base},s}$ and $\mu_{m,s}$ are mean ATE RMSE values over repeated runs. Positive values indicate lower trajectory error than the baseline. Repeatability is summarised by the coefficient of variation

$$\text{CV}(m, s) = \frac{\sigma_{m,s}}{\mu_{m,s}}, \quad (18)$$

where $\sigma_{m,s}$ is the standard deviation of ATE RMSE across the 10 runs.

The evaluation also records coverage, tracking-loss count, relocalisations, keyframes, map points, and tracking time, which are used to check whether trajectory improvement is accompanied by stable tracking behaviour.

4. Results

4.1. Benchmark-level trajectory accuracy

Table 2 summarises benchmark-level trajectory accuracy and repeated-run variability. The baseline mean ATE is 0.071 m for Machine Hall and 0.191 m for Vicon Room. W2 achieves the largest mean ATE improvement, 26.1%, and reduces ATE on 8 of 11 sequences. W1 achieves a mean improvement of 19.7% and the lowest median coefficient of variation, 0.072.

All five variants reduce the benchmark-level mean ATE, while the sequence-level changes remain non-uniform.

4.2. Machine Hall sequences

Figure 2 shows small differences on MH01–MH03. The baseline ATE is approximately 0.037 m on MH01 and 0.032 m on MH02. On MH04, all variants reduce the baseline ATE of 0.120 m, with W3 and W5 reaching approximately 0.055 m. On MH05, W1 achieves the lowest mean ATE, 0.062 m compared with 0.130 m for the baseline.

4.3. Vicon Room sequences

Figure 3 shows ATE reductions on V101, V103, V201, and V202. W1 reduces ATE from 0.119 m to 0.088 m on V101 and from 0.260 m to 0.062 m on V201. W3 reduces ATE from 0.405 m to 0.086 m on V103, while W2 reduces ATE from 0.141 m to 0.060 m on V202. On V102, all variants remain close to the baseline value of 0.064 m.

V203 is the only sequence on which all preprocessing variants increase ATE. Coverage decreases from 0.871 to 0.845–0.856. W2 increases ATE from 0.154 m to 0.169 m, while W3, W4, W5, and W1 reach 0.187 m, 0.292 m, 0.294 m, and 0.306 m, respectively.

4.4. Repeatability and tracking behaviour

Figure 4 shows that W2 achieves the largest average ATE reduction, whereas W1 achieves the lowest median ATE coefficient of variation.

Table 3 shows similar coverage and tracking time for W1, W2, W3, and W5 relative to the baseline. W4 has the lowest coverage, the highest tracking-loss count, and the largest number of keyframes. W5 has the lowest tracking-loss count, while W3 has the lowest number of relocalisations.

4.5. Computational performance

Table 4 shows that W1–W4 require 15.5–19.9 ms per frame, corresponding to preprocessing throughputs of 50–65 fps, whereas W5 is more computationally demanding and reaches 28 fps. For a fixed wavelet basis and decomposition depth, the discrete wavelet transform, coefficient modification, and inverse transform have approximately linear complexity with respect to the number of image pixels, i.e., $O(N)$. In implementations requiring the ordering of coefficient magnitudes, data-dependent SURE threshold selection can introduce an additional cost approaching $O(N \log N)$, which

Table 2

Benchmark-level trajectory accuracy and repeatability of the monocular baseline and the five wavelet preprocessing variants.

Method	MH mean ATE [m]	Vicon mean ATE [m]	Mean ATE imp. vs. Mono [%]	Median CV
Mono (baseline)	0.071	0.191	0.0 (ref.)	0.806
W1	0.049	0.114	19.7	0.072
W2	0.048	0.102	26.1	0.242
W3	0.065	0.128	14.2	0.103
W4	0.055	0.158	11.2	0.271
W5	0.049	0.151	10.5	0.112

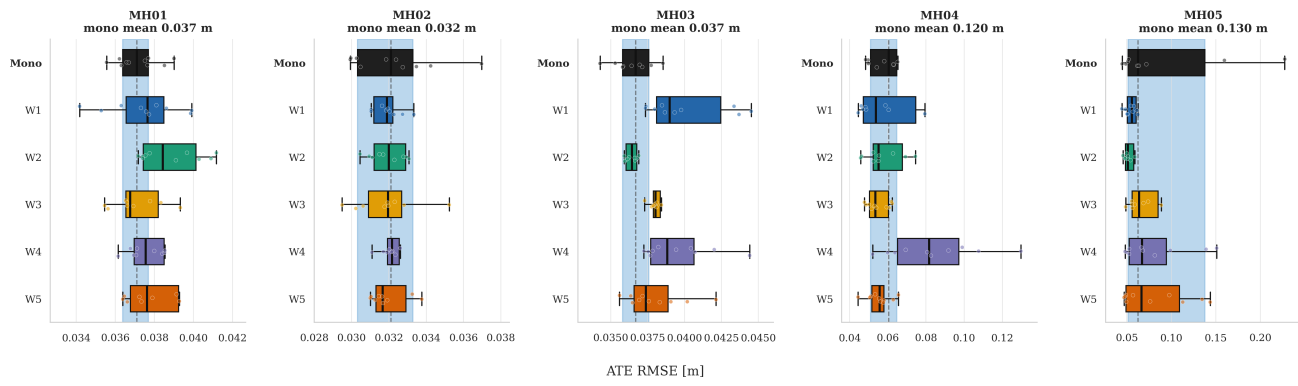


Fig. 2. Machine Hall sequence-wise ATE RMSE over 10 repeated runs for the monocular baseline and the five wavelet preprocessing variants. Method symbols W1–W5 are defined in Table 1.

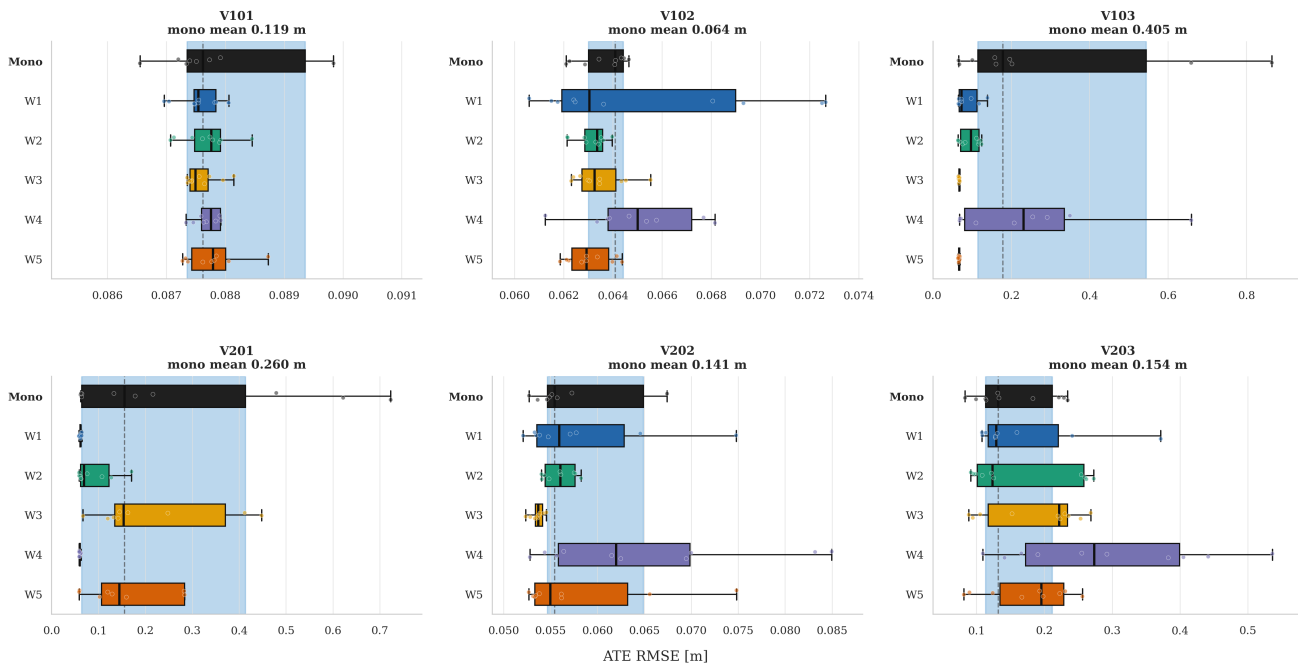


Fig. 3. Vicin Room sequence-wise ATE RMSE over 10 repeated runs for the monocular baseline and the five wavelet preprocessing variants. Method symbols W1–W5 are defined in Table 1.

Table 3

Benchmark-level tracking and map summary for the monocular baseline and the five wavelet variants. Values are averages of sequence-level means over repeated runs.

Method	Mean coverage	Mean tracking losses	Mean relocalisations	Mean keyframes	Mean map points	Mean tracking time [ms]
Mono	0.9444	25.31	0.42	278.8	10923.6	16.92
W1	0.9432	23.87	0.43	284.3	10921.7	15.98
W2	0.9407	24.84	0.35	287.1	10865.4	17.06
W3	0.9423	23.04	0.27	271.5	10738.6	16.77
W4	0.9361	28.11	0.24	313.8	10769.1	16.97
W5	0.9431	21.46	0.40	273.2	10732.9	16.07

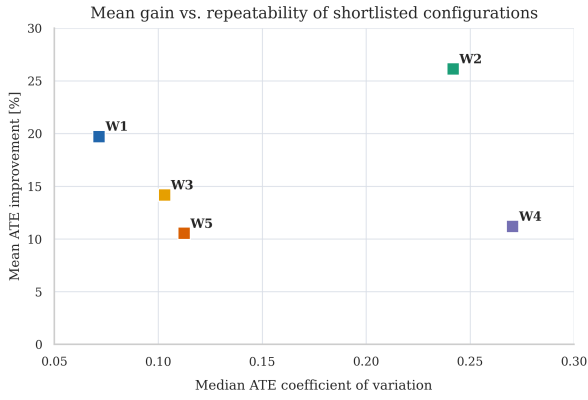


Fig. 4. Trade-off between mean ATE improvement and repeatability. W2 maximises average trajectory improvement, whereas W1 minimises median ATE variation.

Table 4

Single-frame preprocessing time of variants W1–W5 for a 752×480 grayscale image, averaged over 100 repetitions.

Variant	Time [ms]	Std. dev. [ms]	Throughput [fps]
W1	16.3	2.0	62
W2	17.9	3.8	56
W3	19.9	4.2	50
W4	15.5	3.8	65
W5	35.8	6.6	28

is consistent with the higher runtime measured for W5. Combining the separately measured preprocessing times with the mean ORB-SLAM3 tracking times reported in Table 3 gives estimated sequential throughputs of approximately 31.0, 28.6, 27.3, 30.8, and 19.3 fps for W1–W5, respectively. These values are estimates based on sequential execution rather than direct end-to-end measurements.

5. Discussion

The results show that detail-band processing is more effective than approximation-band modification. W2 combines coefficient thresholding with controlled detail amplification and provides the largest mean ATE reduction. W1 suppresses fine-scale fluctuations while reinforcing coarser details and provides the lowest ATE variability. W3 and W5 produce smaller improvements, whereas W4 is associated with reduced coverage, increased tracking losses, and a larger number of keyframes. These observations are consistent with the intensity-comparison mechanisms used by FAST and ORB. Denoising can reduce unstable local contrast variations in (12), while controlled detail amplification can preserve the local intensity ordering represented by (13). This interpretation remains indirect because the recorded data do not include frame-level keypoint-repeatability or descriptor-match statistics.

The superior mean accuracy of W2 suggests that moderate suppression of weak coefficients combined with controlled reinforcement of the remaining detail provides a favourable

balance between noise reduction and structural preservation. W1 modifies the finest fluctuations while reinforcing medium-scale structures, which may explain its lower run-to-run variability. In contrast, W4 changes the approximation component rather than the detail bands and may therefore alter regional brightness without directly improving the repeatability of corners and local descriptors.

The effectiveness of preprocessing depends on the sequence conditions. Larger improvements occur on MH04, MH05, V101, V103, V201, and V202, whereas only minor changes are observed when the baseline error is already low. On V203, rapid motion, viewpoint changes, motion blur, and reduced image overlap coincide with lower coverage and higher ATE for all variants. Fixed wavelet preprocessing is therefore most effective when image quality limits feature extraction and less effective when tracking is dominated by motion dynamics and observation geometry.

6. Conclusions

This study demonstrated that wavelet-domain preprocessing can improve monocular ORB-SLAM3 trajectory estimation by modifying the scale distribution of image information before feature extraction. The evaluation covered five preprocessing variants, all 11 EuRoC MAV sequences, and 660 repeated runs. W2 achieved the largest mean ATE improvement of 26.1% and reduced trajectory error on 8 of 11 sequences, whereas W1 provided the highest repeatability, with a median ATE coefficient of variation of 0.072. The results also showed that detail-band conditioning is more effective than approximation-band modification for preserving image structures relevant to FAST detection and ORB description.

A natural extension of this work is an adaptive preprocessing module that selects the wavelet family, decomposition depth, thresholding strategy, and detail gain according to the current image and operating environment. Such a module could use online estimates of illumination, contrast, noise level, motion blur, texture density, and tracking quality to adjust image conditioning before feature extraction. This direction enables the development of adaptive visual SLAM systems that respond to changing lighting, scene structure, and platform dynamics while maintaining stable trajectory estimation across diverse operating conditions [28].

Future work will evaluate whether the observed effects generalise to other feature-based, direct, semi-direct, and visual-inertial SLAM algorithms. Curvelet-based preprocessing will also be investigated because its directional multiscale representation may improve the preservation of elongated edges and curved structures. The evaluation will additionally be extended to indoor and outdoor datasets containing stronger illumination changes, low-texture scenes, motion blur, and adverse environmental conditions, as well as to end-to-end real-time tests on embedded computing platforms.

Acknowledgements

This work was supported by the Military University of Technology, Poland, under research project UGB 127/2026.

References

- [1] Neyestani, A., Picariello, F., Ahmed, I., Daponte, P. & Vito, L. D. From pixels to precision: A survey of monocular visual odometry in digital twin applications. *Sensors* **24**, 1274 (2024). <https://doi.org/10.3390/s24041274>.
- [2] Kowalski, J. K. & Chodnicki, M. Direct optical path measurement in fibre-optic gyroscopes: A potential method for compensating slow-drifting errors. *Opto-Electron. Rev.* **33**, e154748 (2025). <https://doi.org/10.24425/opelre.2025.154748>.
- [3] Pasek, P. & Kaniewski, P. Energy-efficient distributed estimation algorithm for wireless sensor networks based on covariance intersection with eigendecomposition. *Metrol. Meas. Syst.* **31**, 465–480 (2024). <https://doi.org/10.24425/mms.2024.150290>.
- [4] Chen, L. *et al.* A survey of computer vision detection, visual SLAM algorithms, and their applications in energy-efficient autonomous systems. *Energies* **17**, 5177 (2024). <https://doi.org/10.3390/en17205177>.
- [5] Stawowy, M., Paś, J., Klimczak, T., Rosiński, A. & Duer, S. W. Modeling uncertainty in estimating the information quality acquired from the fire alarm systems. *Inf. Sci.* **736**, 123102 (2026). <https://doi.org/10.1016/j.ins.2026.123102>.
- [6] Campos, C., Elvira, R., Rodriguez, J. J. G., Montiel, J. M. M. & Tardos, J. D. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap SLAM. *IEEE Trans. Robot.* **37**, 1874–1890 (2021). <https://doi.org/10.1109/TRO.2021.3075644>.
- [7] Mur-Artal, R. & Tardos, J. D. ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras. *IEEE Trans. Robot.* **33**, 1255–1262 (2017). <https://doi.org/10.1109/TRO.2017.2705103>.
- [8] Rosten, E. & Drummond, T. Machine learning for high-speed corner detection. In *Proceedings of the European Conference on Computer Vision*, 430–443 (Springer, 2006). https://doi.org/10.1007/11744023_34.
- [9] Calonder, M., Lepetit, V., Strecha, C. & Fua, P. BRIEF: Binary robust independent elementary features. In *European Conference on Computer Vision*, 778–792 (Springer, 2010). https://doi.org/10.1007/978-3-642-15561-1_56.
- [10] Rublee, E., Rabaud, V., Konolige, K. & Bradski, G. ORB: An efficient alternative to sift or surf. In *Proceedings of the IEEE International Conference on Computer Vision*, 2564–2571 (IEEE, 2011). <https://doi.org/10.1109/ICCV.2011.6126544>.
- [11] Engel, J., Schoeps, T. & Cremers, D. LSD-SLAM: Large-scale direct monocular SLAM. In *European Conference on Computer Vision*, 834–849 (Springer, 2014). https://doi.org/10.1007/978-3-319-10605-2_54.
- [12] Engel, J., Koltun, V. & Cremers, D. Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 611–625 (2018). <https://doi.org/10.1109/TPAMI.2017.2658577>.
- [13] Qin, T., Li, P. & Shen, S. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Trans. Robot.* **34**, 1004–1020 (2018). <https://doi.org/10.1109/TRO.2018.2853729>.
- [14] Yu, L., Yang, E. & Yang, B. AFE-ORB-SLAM: Robust monocular VSLAM based on adaptive fast threshold and image enhancement for complex lighting environments. *J. Intell. Robot. Syst.* **105**, 26 (2022). <https://doi.org/10.1007/s10846-022-01645-w>.
- [15] Chen, J., Wang, Y., Hou, P., Chen, X. & Shao, Y. Dark-SLAM: A robust visual simultaneous localization and mapping pipeline for an unmanned driving vehicle in a dark night environment. *Drones* **8**, 390 (2024). <https://doi.org/10.3390/drones8080390>.
- [16] Qiu, H. *et al.* An improved underwater visual SLAM through image enhancement and sonar fusion. *Remote Sens.* **16**, 2512 (2024). <https://doi.org/10.3390/rs16142512>.
- [17] Wang, Z., Cheng, Q. & Mu, X. RU-SLAM: A robust deep-learning visual simultaneous localization and mapping (SLAM) system for weakly textured underwater environments. *Sensors* **24**, 1937 (2024). <https://doi.org/10.3390/s24061937>.
- [18] Pretto, A., Menegatti, E., Jitsukawa, Y., Ueda, R. & Arai, T. Image similarity based on discrete wavelet transform for robots with low-computational resources. *Robot. Auton. Syst.* **58**, 879–888 (2010). <https://doi.org/10.1016/j.robot.2010.03.009>.
- [19] Yu, H., Zhu, H. & Huang, F. Visual simultaneous localization and mapping (SLAM) based on blurred image detection. *J. Intell. Robot. Syst.* **103**, 12 (2021). <https://doi.org/10.1007/s10846-021-01456-5>.
- [20] Lange, M., Raisch, C. & Schilling, A. WLD: A wavelet and learning based line descriptor for line feature matching. In *Vision, Modeling, and Visualization* (The Eurographics Association, 2020). <https://doi.org/10.2312/vmv.20201186>.
- [21] Mallat, S. G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693 (1989). <https://doi.org/10.1109/34.192463>.
- [22] Donoho, D. L. & Johnstone, I. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455 (1994). <https://doi.org/10.1093/biomet/81.3.425>.
- [23] Donoho, D. L. & Johnstone, I. M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **90**, 1200–1224 (1995). <https://doi.org/10.1080/01621459.1995.10476626>.
- [24] Chang, S. G., Yu, B. & Vetterli, M. Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Process.* **9**, 1532–1546 (2000). <https://doi.org/10.1109/83.862633>.
- [25] Mihcak, M. K., Kozintsev, I., Ramchandran, K. & Moulin, P. Low-complexity image denoising based on statistical modeling of wavelet coefficients. *IEEE Signal Process. Lett.* **6**, 300–303 (1999). <https://doi.org/10.1109/97.803428>.
- [26] Burri, M. *et al.* The EuRoC micro aerial vehicle datasets. *Int. J. Robot. Res.* **35**, 1157–1163 (2016). <https://doi.org/10.1177/0278364915620033>.
- [27] Kraszewski, T., Panasiuk, J., Słowak, P. & Kaniewski, P. Testing positioning system for ground penetrating radar using multi-degree-of-freedom industrial robot. *Metrol. Meas. Syst.* **32**, 1–15 (2025). <https://doi.org/10.24425/mms.2025.154681>.
- [28] Krzykowska-Piotrowska, K., Rosiński, A., Paś, J., Piotrowski, M. & Siergiejczyk, M. Optimizing the operational process of a social robot for elderly assistance: Enhancing reliability and readiness. *Electronics* **14**, 1630 (2025). <https://doi.org/10.3390/electronics14081630>.