Electronics and Information Technologies, 2026, 33, 113–130
http://publications.lnu.edu.ua/collections/index.php/electronics/index

UDC: 004.932.2

SPATIAL-GEOMETRIC EVALUATION OF LOCAL FEATURES IN MONOCULAR VISUAL ODOMETRY

Andriy Fesiuk**, Yuriy Furgala
Department of Optoelectronics and Information Technologies
Ivan Franko National University of Lviv,
50 Drahomanova St., 79005 Lviv, Ukraine

Fesiuk A. V., Furgala Y. M. (2026). Spatial-Geometric Evaluation of Local Features in Monocular Visual Odometry. *Electronics and Information Technologies*, 33, 113–130. https://doi.org/10.30970/eli.33.9

ABSTRACT

Background. Monocular visual odometry is an important component of visual navigation systems. However, its accuracy depends on the quality of local features and inter-frame correspondences. In the VO task, not only is geometric consistency important, but also motion observability, the physical validity of the recovered configuration, and the spatial-structural properties of local features. This study aims to provide a comprehensive evaluation of keypoint detection and description methods for monocular visual odometry.

Materials and Methods. The study was conducted on the EuRoC MAV dataset. The ORB, BRISK, AKAZE, KAZE, SIFT, SURF, and SuperPoint methods were analyzed for the number of keypoints, ranging from 200 to 1000. Motion estimation was performed using the essential matrix, the USAC_FAST filter, the recoverPose method, a minimum parallax check, and spatially guided keypoint selection. The accuracy of the recovered trajectory was evaluated using the APE and RPE metrics. To analyze the quality of local features and correspondences, the geometric component, the parallax indicator, the correct cheirality ratio, and metrics of keypoint coverage uniformity, local redundancy, and structural consistency were used. An integral quality indicator was applied to summarize the results.

Results and Discussion. The geometric metrics most often highlight AKAZE and SURF, whereas SuperPoint shows strong performance in terms of spatial characteristics. In terms of the structural consistency of correspondences, SURF consistently demonstrates the best results. As the number of keypoints increases, most methods show an initial improvement followed by saturation, and in some cases, a deterioration of individual characteristics. SURF was found to be the most balanced method across the set of criteria, whereas ORB showed the weakest results in most cases. The correlation analysis showed that the informativeness of the metrics varies by sequence type.

Conclusion. The proposed approach confirmed the relevance of multicriteria evaluation of local features in monocular visual odometry. It was shown that no single metric is universal across all scene types. In contrast, the integral indicator enables the summary of different aspects of quality and a more well-grounded ranking of the methods.

Keywords: monocular visual odometry, keypoint detection, image matching, motion estimation, deep learning, neural networks.

INTRODUCTION

Visual navigation and robotics systems widely employ monocular visual odometry as one of the approaches for estimating camera motion and reconstructing the motion trajectory. Under real operating conditions, the performance of such systems is

complicated by dynamic viewpoint changes, illumination variations, motion blur, and the presence of weakly textured scene regions. The accuracy of odometric estimates depends on the quality of correspondences established between neighboring frames, since these correspondences form the basis for determining the relative camera pose. Errors in keypoint matching, instability of the correspondence set, and insufficient observability of inter-frame displacement adversely affect motion estimation accuracy, accumulate over time, and increase trajectory error [1-5].

Keypoints and their descriptors, which are considered in this work as local features, constitute the foundation of monocular visual odometry. In most approaches to local feature analysis, the primary focus is placed on geometric characteristics, in particular, matching accuracy, the inlier ratio, detection repeatability, and descriptor stability [5, 6]. Although these indicators are important, in the task of monocular visual odometry, they do not provide a complete characterization of feature suitability, because correspondence correctness within a geometric model does not guarantee reliable motion recovery. In particular, under small parallax or near-degenerate, geometrically weak configurations, even formally correct correspondences may lead to unstable translation estimation [7]. It is also important to assess how consistent the recovered geometry is at the current step, specifically whether the correctness of depth signs is preserved during camera pose recovery. In addition, the spatial-structural properties of inliers are of considerable importance: non-uniform image coverage, excessive local concentration of points, or weak correspondence to the scene structure reduce the reliability of motion estimation and increase sensitivity to noise [1-4].

Existing studies on visual odometry mainly examine either final trajectory errors or individual characteristics of local features and correspondences [8-11]. While this analysis is helpful, it does not fully reveal the connection between feature properties and the odometry algorithm's actual performance. Previous research [6] showed that for image matching, it is important to consider not only geometric consistency but also the spatial structure of keypoints. However, directly applying this method to monocular visual odometry is inadequate because it also requires accounting for motion observability, the stability of camera pose recovery, and the link between local feature characteristics and trajectory-based metrics.

Modern approaches to image keypoint detection can be divided into classical methods, including SIFT [12], SURF [13], KAZE [14], AKAZE [15], ORB [16], and BRISK [17], and deep learning-based solutions such as SuperPoint [18]. Classical algorithms demonstrate high computational efficiency and mathematical interpretability; however, they are often vulnerable to changes in illumination or weak scene texture. At the same time, neural network-based methods are more robust under challenging conditions [8, 19]. Still, their specific nature of point localization and the statistical properties of the inlier distribution require comparative analysis [9, 18]. Evaluating the influence of the local feature type on geometric reliability and monocular visual odometry errors remains an important task, since different methods form keypoint representations in different ways, which in turn affects the stability of the navigation system.

The goal of this work is to analyze keypoint detection and description methods for monocular visual odometry using a comprehensive quality assessment approach. For this purpose, an integral index is employed that combines the geometric consistency of correspondences, indicators of motion observability and camera pose recovery stability, and the structural-spatial properties of consistent correspondences.

MATERIALS AND METHODS

For the experimental validation of the proposed approach, the EuRoC MAV dataset [20] was used. It includes scenes with varying levels of motion complexity, spatial structure, and visual conditions, enabling the evaluation of local features across easy, medium, and

difficult monocular visual odometry scenarios. For the analysis, images from the cam0 camera were processed. Two groups of sequences were considered:

- Machine Hall - large industrial indoor environments with non-uniform illumination and a considerable number of repetitive structures. Three difficulty levels were investigated: MH_01_easy, MH_03_medium, and MH_05_difficult. In this group, the difficulty increases from relatively smooth motion and more favorable illumination conditions to faster movement under dim lighting, which may be accompanied by stronger motion blur and reduced feature contrast [20].
- Vicon Room - a confined indoor environment with highly accurate ground-truth motion. Although this setting is more controlled, the increasing difficulty in this group is mainly associated with camera dynamics and illumination conditions. Three sequences with different difficulty levels were used: V1_01_easy, V1_02_medium, and V1_03_difficult. The transition to more challenging sequences is characterized by sharper maneuvers, faster viewpoint changes, and temporary loss of textured objects from the field of view, which makes tracking and motion estimation more difficult [20].

To compare the quality of local features in visual odometry, the classical methods ORB, AKAZE, KAZE, SIFT, SURF, and BRISK, as well as the deep learning-based method SuperPoint, were investigated. In this work, an off-the-shelf SuperPoint implementation from the LightGlue library [21] was used without any additional model retraining.

The analysis was performed for different numbers of selected keypoints, ranging from 200 to 1000 in increments of 200. Examples of the images used in the experiments are shown in **Fig. 1**.

To reduce the effect of local keypoint concentration in the most contrast-rich image regions, a spatially guided selection strategy based on a regular 8×5 grid was applied. This approach provides more balanced image coverage, since even with a large number of correct correspondences, their spatial clustering may degrade the stability of scene geometry estimation, reduce reconstruction reliability, and lead to instability of the recovered motion [22, 23].

Motion estimation was performed in a calibrated monocular setting based on the essential matrix [24]. For each pair of frames with a stride = 2, a set of keypoint correspondences was established using Brute-Force matching, followed by filtering according to the Lowe ratio criterion with a threshold of 0.75 [12]. Before geometric estimation, point

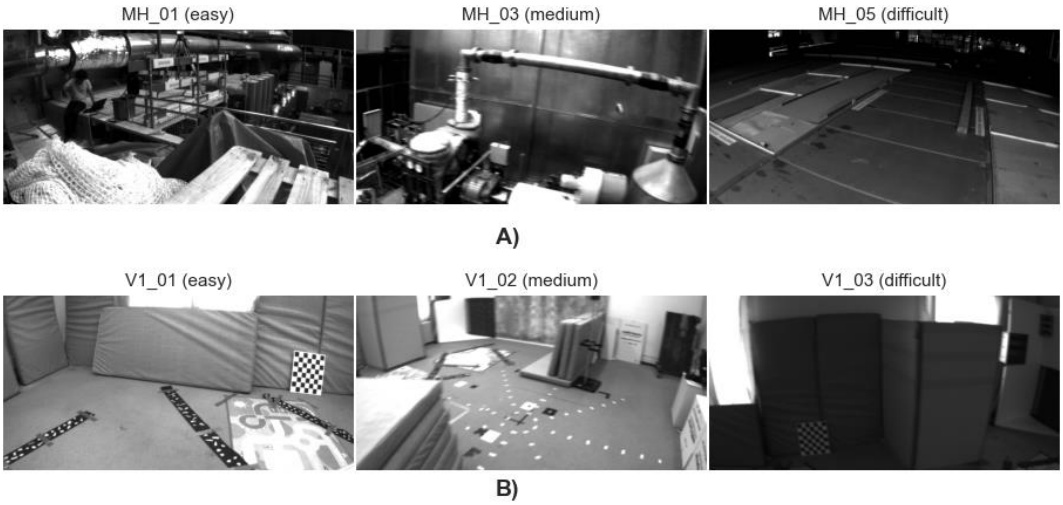

Fig. 1. Examples of images from the EuRoC MAV dataset used in the experiments: A) frames from the Machine Hall group (MH_01_easy, MH_03_medium, MH_05_difficult); B) frames from the Vicon Room group (V1_01_easy, V1_02_medium, V1_03_difficult).

coordinates were corrected using the camera calibration parameters to remove distortion effects, thereby reducing systematic errors in correspondence localization [24].

The essential matrix was estimated using USAC_FAST with a threshold of 1.5 px, a confidence level of 0.999, and a maximum of 5000 iterations [25, 26]. To ensure robust estimation, additional conditions were imposed: at least 50 initial correspondences and at least 30 inliers after geometric verification. The relative pose between frames was recovered using the recoverPose procedure, which selects the physically consistent decomposition of the essential matrix and returns a mask of points with positive depth. To reduce the influence of weakly observable configurations, an additional minimum parallax check of 5 px was applied in the freeze_translation mode, in which the translation estimate was not updated when the inter-frame displacement was insufficient [7, 27].

The quality of the recovered trajectory was evaluated using the evo package, a standard tool for trajectory analysis in robotics, SLAM, and visual odometry. For external accuracy assessment, the APE (Absolute Pose Error) and RPE (Relative Pose Error) metrics were used [28, 29].

The APE metric was used to assess the global trajectory error after alignment of the estimated and reference trajectories. In this work, Sim(3) alignment was employed, that is, alignment accounting for rotation, translation, and a single global scale factor, which is particularly important for monocular visual odometry, where the absolute scale cannot be determined uniquely without additional sensors. Thus, APE characterizes the accumulated drift and the overall deviation of the estimated trajectory from the reference in a global sense [29].

The RPE metric was used to assess the local error of relative motion over a fixed temporal baseline. Unlike APE, it reflects the stability of short-term estimation and is less sensitive to long-term error accumulation. To ensure comparability across all runs, the analysis was performed over a consistent temporal interval of 0.30 s, which was converted into the corresponding number of frames, accounting for the sampling stride [28]. At a frame rate of 20 Hz, this corresponds to a baseline interval of approximately 6 frames.

Since the translational RPE value in metric units strongly depends on the amplitude of inter-frame displacement, this work additionally considered a normalized form of the relative error:

$$RPE_{norm} = \frac{RPE_t}{\varepsilon + d_{med}}, \tag{1}$$

where RPE_t denotes the aggregated translational relative error, d is the median inter-frame displacement of the corresponding points in pixels, and ε is a small positive stabilizing term. Such normalization reduces the influence of varying motion intensity across individual fragments and enables a more meaningful comparison of cases with similar absolute error but different motion observability. In addition, for the analysis of rotational stability, tail-risk indicators derived from the angular RPE were used, in particular rotation error percentiles and the proportions of large deviations.

The internal quality assessment of local features in the VO task is based on a combination of geometric and spatial-structural components.

The basic geometric characteristic of a correspondence set is the proportion of inliers among all verified matches. It is calculated as the ratio of inliers consistent with the estimated geometric model to the total number of correspondences [30]. However, the inlier ratio alone does not indicate how closely these correspondences match the estimated model. Two correspondence sets may have similar inlier ratio values but differ significantly in residual geometric error. Therefore, a modified indicator is used in this work:

$$G_{ratio} = \frac{N_{inl}}{N_{match}} exp(-\overline{e}), \tag{2}$$

$$\overline{e} = \frac{med(e^2_{Sampson})}{\tau^2}, \tag{3}$$

where $med(e^2_{Sampson})$ is the median squared Sampson error for the inliers [27], and τ is the geometric threshold used in the USAC verification procedure. High values of Gratio correspond to cases in which the correspondence set simultaneously exhibits a high inlier ratio and a low residual geometric error.

For reliable motion recovery, geometric consistency of correspondences alone is not sufficient, since the quality of translation estimation strongly depends on the magnitude of inter-frame point displacement. When the displacement is small, the motion becomes less informative, and the recovery of spatial structure becomes less stable [7, 30]. Therefore, an additional parallax-based indicator is used:

$$P_{score} = 1 - \exp\left(-\frac{d_{med}}{p_0}\right), \tag{4}$$

where d_{med} is the median inter-frame displacement of corresponding points in pixels, and p_0 is a reference parameter. This function increases with parallax: for small displacement, P_{score} it approaches zero, whereas for sufficiently large displacement, it approaches one. Therefore, the indicator P_{score} characterizes motion observability and the suitability of the current configuration for stable estimation of translation [27, 30].

The second characteristic is the cheirality indicator:

$$Cheirality_{ratio} = \frac{N_{chier}}{N_{inl}}, \tag{5}$$

where N_{chier} is the number of inliers for which positive depth is obtained after pose recovery, and N_{inl} is the total number of inliers. This indicator follows from the cheirality check: a physically valid solution is the one for which the reconstructed 3D points lie in front of both cameras [24, 27]. Therefore, $Cheirality_{ratio}$ characterizes the physical validity of the recovered configuration. High values of $Cheirality_{ratio}$ correspond to cases in which most inliers support a correct spatial solution, whereas a decrease in this indicator may indicate estimation instability or a weak 3D interpretation [30].

Taking these factors into account, the final geometric component for the visual odometry task was defined as follows:

$$G_{VO} = G_{ratio} \cdot P_{scoe} \cdot Cheirality_{ratio}. \tag{6}$$

This form accounts for the fact that, for reliable motion estimation, a high inlier ratio alone isn't enough. Adequate inter-frame displacement and the physical correctness of the recovered spatial structure are also essential.

The spatial-structural component is based on three metrics proposed in the previous study [6]: CUI, SCS, and RI. The Coverage Uniformity Index (CUI) evaluates the uniformity of image-plane coverage. The Scene Consistency Score (SCS) measures the consistency between the scene structure and the structural composition of the point set.

To assess the local spatial redundancy of inliers, a modified Redundancy Index (RI) was used. Unlike the basic version, which evaluates redundancy using a fixed neighborhood threshold [6], this work defines the neighborhood radius relative to the image diagonal. Such normalization reduces the metric's dependence on image resolution and enables a more meaningful comparison of configurations with different point counts. For

each point, the number of neighbors within a radius $r = \gamma D$ is counted, where D is the image diagonal. The obtained local neighbor count is then compared with the expected density level for the current point set [31-33]. The final RI value is defined as the average over all points. It takes values in the range [0,1]: low values correspond to a more uniform distribution, whereas high values indicate local clustering and point redundancy.

This modification follows from the fact that, for local feature analysis tasks, it is important to consider not only the total number of points but also the nature of their spatial arrangement. Previous studies [6] have shown that the spatial distribution of keypoints is an independent quality characteristic, since excessive local concentration reduces scene coverage even when a large number of features is available [33]. In this work, this idea is further developed through diagonal normalization of the radius and explicit consideration of the expected local density, allowing the RI indicator to better reflect redundancy itself rather than the absolute point density.

For the integral description of spatial-structural quality, the following component was used:

$$S_{VO} = \frac{1}{3}(CUI + (1 - RI)^2 + SCS), \tag{7}$$

where the quadratic term $(1 - RI)^2$ increases the penalty for local redundancy. The use of $(1 - RI)^2$, rather than the linear form $(1 - RI)$, makes it possible to distinguish more clearly between point sets with moderate and high clustering.

The final integral indicator for the monocular visual odometry task is defined as:

$$Q_{VO} = 0.62 \cdot G_{VO} + 0.38 \cdot S_{VO}, \tag{8}$$

where the geometric component is assigned a higher weight, since it is the primary factor determining the correctness of motion recovery, whereas the spatial-structural component explains the stability of estimation and robustness to local deformations and non-uniform scene structure.

Since visual odometry is a stepwise procedure and may contain failed steps, an additional penalty was introduced for the integral score based on the frequency of unsuccessful steps. Let $f \in [0,1]$ denote the proportion of failed steps in a run (*fail rate*). Then the penalized score is defined as:

$$Q_{pen} = Q_{VO} \cdot (1 - f). \tag{9}$$

Thus, the proposed methodology combines two levels of evaluation. The external level characterizes the actual trajectory accuracy as measured by APE and RPE. In contrast, the internal level explains this behavior through the geometric reliability of correspondences, motion observability, and the spatial-structural balance of inliers. Such a design makes it possible not only to rank the methods, but also to interpret the reasons for their behavior under monocular visual odometry conditions.

RESULTS AND DISCUSSION

In all the presented plots, the values of the analyzed metrics are reported as the median for each experiment. This form of representation was chosen because the distributions of indicators in the monocular visual odometry task may contain outliers and locally unstable estimates. In contrast, the median provides a more robust and representative summary.

Fig. 2 shows the dependence of the geometric consistency index, G_{ratio}, on the number of keypoints, N. In general, AKAZE demonstrates the highest metric values in most sequences, while SURF is usually second or close to the best result. For AKAZE, the G_{ratio} values remain consistently high: in the MH_01_easy, V1_01_easy, and MH_05_difficult sequences, they are close to 0.91-0.93, while in the more challenging V1_03_difficult sequence, they remain at approximately 0.82.

SIFT and KAZE also show fairly high results; however, they generally remain below AKAZE and SURF. For example, at N=1000 in V1_01_easy, the G_{ratio} value is 0.929 for AKAZE, 0.905 for SURF, 0.903 for SIFT, and 0.884 for KAZE.

ORB and BRISK provide lower-level geometric consistency. In most sequences, ORB yields the lowest values among the classical methods, while BRISK usually outperforms ORB but remains noticeably inferior to AKAZE, SURF, and SIFT. For example, in MH_03_medium at N=1000, the value of G_{ratio} for BRISK equals 0.784, whereas for AKAZE it reaches 0.905, corresponding to a difference of approximately 15.4%.

SuperPoint demonstrates the most pronounced negative trend: as N increases, its G_{ratio} value systematically decreases in all sequences. In particular, in MH_01_easy, the indicator decreases from 0.776 to 0.693 (approximately 10.7%); in V1_01_easy, from 0.691 to 0.609 (11.9%); and in V1_03_difficult, from 0.576 to 0.487 (15.5%). This indicates that increasing the number of points for this method does not improve geometric consistency but, on the contrary, worsens it. However, in some cases, SuperPoint was more effective than ORB. It can also be observed that SuperPoint 's performance depends on sequence difficulty. For classical methods, the dependence of performance on sequence difficulty is less evident.

Thus, according to the G_{ratio} metric, the most stable results were achieved by AKAZE and SURF, whereas SuperPoint proved most sensitive to increasing the number of points. Overall, the obtained results confirm that increasing the number of keypoints by itself does not guarantee an improvement in the geometric quality of correspondences.

Fig. 3 presents the dependence of the parallax indicator P_{score} on the number of keypoints. Unlike the G_{ratio} metric, the separation between methods is considerably weaker in this case, and in some sequences, the values from different detectors remain very close. This indicates that, within the present experiment, the parallax indicator should rather be regarded as an indicator of overall motion observability than as an independent criterion for clear method ranking.

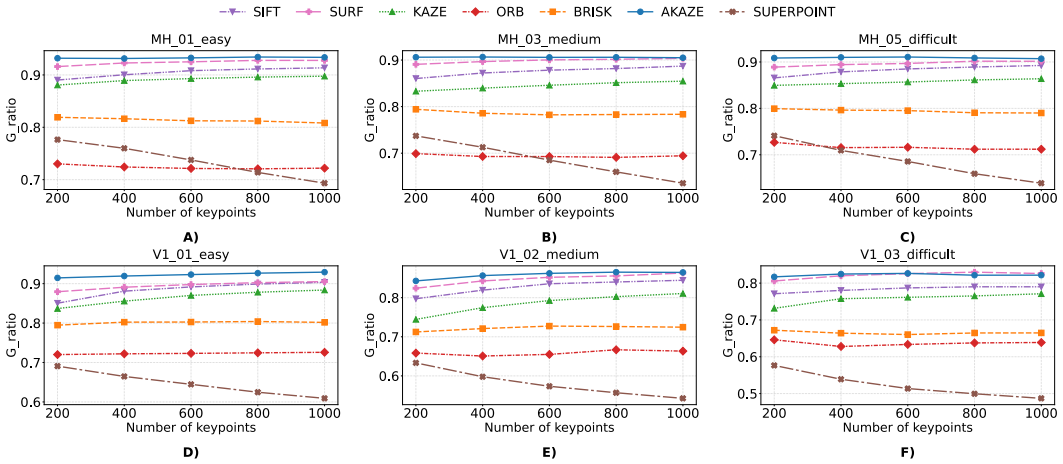

Fig. 2. Median values of G_{ratio} vs. the number of keypoints for different detectors in the EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

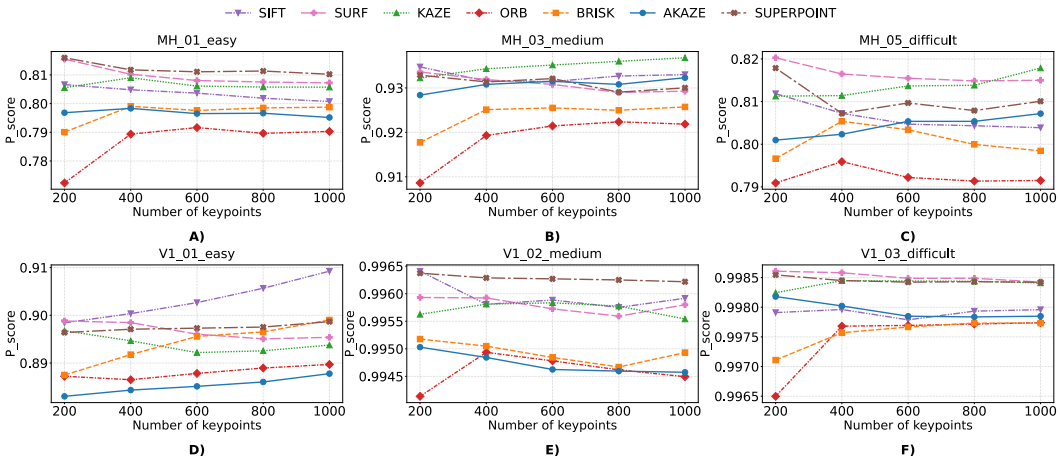

Fig. 3. Median values of P_{score} vs. the number of keypoints for different detectors in the EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

The most noticeable separation between detectors is observed in the MH_01_easy, MH_03_medium, and MH_05_difficult sequences, where certain methods produce higher or lower P_{score} values, although the overall difference remains moderate. In these cases, ORB is more often found in the lower part of the group, whereas SIFT, SURF, KAZE, or SuperPoint yield slightly better results depending on the sequence. At the same time, as the number of points increases, most methods exhibit either a slight improvement or a rapid transition to a saturation.

A different pattern is observed for V1_02_medium and V1_03_difficult, where almost all methods yield very high, closely matched P_{score} values. Under such conditions, this metric does not effectively separate the detectors, indicating its limited discriminative ability in sequences where motion observability is generally favorable for most methods.

Thus, P_{score} should primarily be considered an auxiliary geometric indicator that characterizes the conditions for reliable motion estimation but is not by itself sufficient for the final comparison of detectors. For this reason, its interpretation is most meaningful when combined with G_{ratio}, the proportion of correct cheirality, and the integral quality indicator.

Fig. 4 illustrates the dependence of the correct cheirality ratio on the number of keypoints. The obtained results indicate that this metric depends more strongly on the sequence type than P_{score}: across different scenes, not only the absolute levels of the values change, but also the relative positions of the methods. This means that a complex interaction between detector properties and the characteristics of a particular sequence determines the physical validity of the recovered spatial configuration.

Unlike P_{score}, which in some cases hardly separates the methods, the correct cheirality ratio more often reveals pronounced differences between detectors. At the same time, no universal leader is observed: in different sequences, the best results are demonstrated by SURF, SIFT, KAZE, AKAZE, or SuperPoint. This pattern is indicative, as it shows that even under similar motion observability conditions, different methods may differ substantially in their ability to form correspondences suitable for physically valid scene geometry recovery.

Therefore, this metric is an important complement to the other geometric indicators, as it allows evaluation not only of the consistency of correspondences but also of their suitability for correct spatial reconstruction. At the same time, its results further confirm the relevance of a comprehensive analysis, because it does not, on its own, provide a complete ranking of the methods.

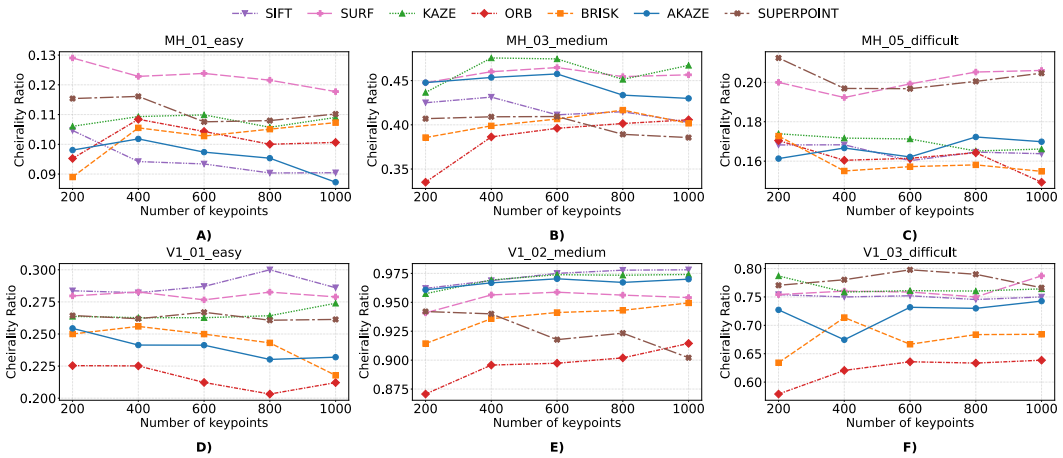

Fig. 4. Median values of Cheirality ratio vs. the number of keypoints for different detectors in the EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

Fig. 5 shows the dependence of the inlier coverage uniformity indicator CUI on the number of keypoints. A common trend is observed across all investigated sequences: as the number of keypoints increases, the metric value either rises or gradually reaches a saturation regime. This indicates that increasing the number of keypoints generally contributes to a more complete and more uniform spatial coverage of the scene by correct correspondences.

SuperPoint primarily demonstrates the highest CUI values in all six sequences, while SURF is usually the second-best method. For example, in MH_01_easy at N=1000, the CUI value for SuperPoint is approximately 0.806, whereas for SURF it is 0.778, for AKAZE 0.706, and for ORB 0.598. Thus, the advantage of SuperPoint over ORB is approximately 34.8%, and over AKAZE about 14.2%. A similar pattern is also observed for MH_03_medium, where at N=1000 SuperPoint reaches approximately 0.756, whereas ORB achieves only 0.555.

In V1_01_easy and V1_03_difficult, the absolute metric values are lower across all methods; however, the relative separation between them remains. For example, in

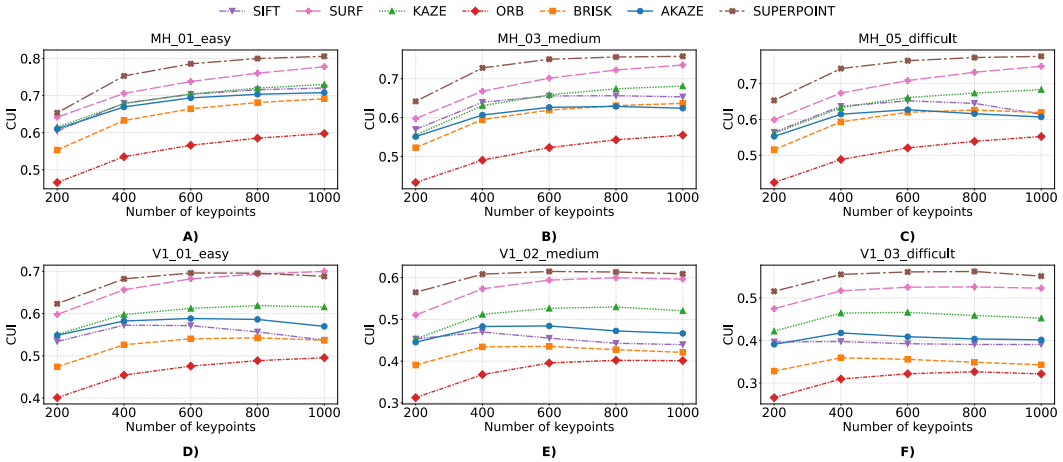

Fig. 5. Median values of CUI vs. the number of keypoints for different detectors in the EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

V1_03_difficult at N=1000, the value for SuperPoint is approximately 0.551, for SURF 0.523, whereas for ORB it is only 0.323. This means that SuperPoint exceeds ORB by approximately 1.7 times. Thus, even under more challenging conditions, SuperPoint and SURF remain in the upper part of the group, while ORB consistently demonstrates the lowest or near-lowest values.

AKAZE, KAZE, and SIFT form a middle group of methods, characterized by either a moderate increase in CUI or stabilization after N=400-600. BRISK also improves coverage as the number of points increases; however, in most cases, it remains inferior to this group. Therefore, the CUI metric effectively reflects the spatial completeness of scene coverage by inliers and shows that SuperPoint provides the most uniform distribution of correct correspondences over the image. At the same time, the absolute level of the metric also depends on the particular sequence, so its values should primarily be interpreted within each scene separately.

Fig. 6 presents the dependence of the normalized indicator of local inlier redundancy, RI, on the number of keypoints, with lower values indicating better results. A common trend is observed across all investigated sequences: as N increases, the RI value rises, that is, local redundancy becomes stronger. This means that as the number of points increases, inliers increasingly concentrate in individual local regions, even as the overall scene coverage improves.

SuperPoint consistently shows the lowest RI values across all six sequences. For example, in MH_01_easy at N=1000, its value is approximately 0.303, whereas for ORB it is 0.835, for BRISK 0.659, and for AKAZE 0.600. Thus, relative to ORB, the value for SuperPoint is approximately 2.8 times lower, and relative to AKAZE, almost 2 times lower. A similar pattern is observed in other sequences as well, in particular in MH_03_medium, MH_05_difficult, and V1_03_difficult.

The worst results are usually shown by ORB, which, in all cases, has the highest or near-highest metric values. For example, in V1_01_easy at N=1000, the ORB value is approximately 0.886, whereas for SIFT it is 0.714, for SURF 0.570, and for SuperPoint 0.455. This indicates a substantially higher local concentration of inliers for ORB compared with the other methods.

In most sequences, SIFT and SURF form a relatively favorable group, with lower redundancy than AKAZE, KAZE, and BRISK. For SIFT and SURF, the increase in RI when moving from N=200 to N=400 is noticeable, but afterwards the changes become smaller or

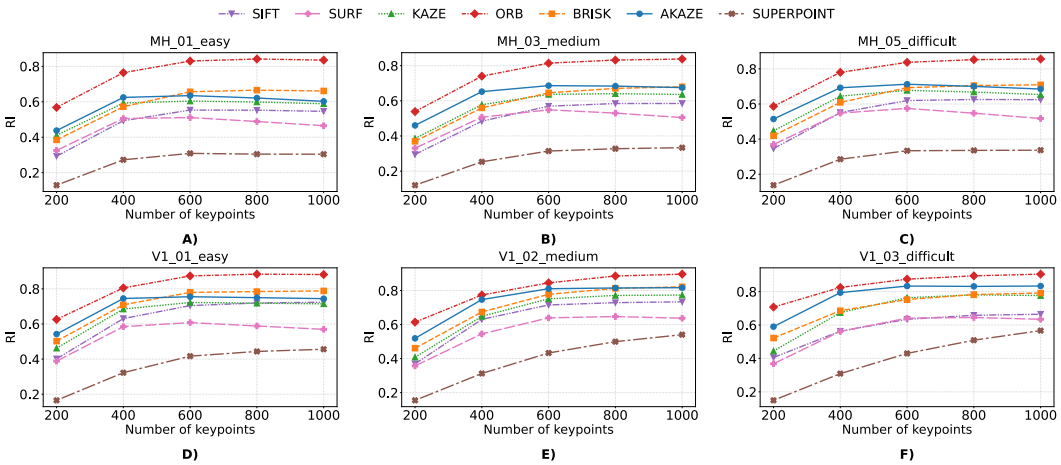

Fig. 6. Median values of RI vs. the number of keypoints for different detectors in the EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

reach a saturation regime. In contrast, for ORB, BRISK, AKAZE, and KAZE, the metric values remain at a higher level in almost all sequences.

Thus, the RI metric complements CUI well: if CUI reflects the completeness of scene coverage by inliers, then RI shows how strongly these inliers are locally clustered. In this respect, SuperPoint proved to be the best, whereas ORB was the weakest. At the same time, the absolute level of the metric partly depends on the particular sequence, so its values should primarily be interpreted within each scene separately.

Fig. 7 illustrates the dependence of the inlier structural consistency indicator SCS on the number of keypoints N. Unlike CUI and RI, a different group of leaders can be clearly identified here: in all investigated sequences, the highest SCS values are consistently demonstrated by SURF, while KAZE and AKAZE usually form a second group of methods that also achieve high results.

For SURF, the metric values are the highest in all six sequences and, in most cases, either increase with the number of keypoints or remain close to their maximum levels. For example, in MH_01_easy at N=1000, the SCS value for SURF is approximately 0.922, whereas for AKAZE it is 0.779, for SIFT 0.494, and for ORB 0.297. Thus, the advantage of SURF over AKAZE is approximately 18.4%, while over ORB it is about threefold. A similar pattern is also observed for V1_03_difficult, where at N=1000 SURF exceeds AKAZE by approximately 18.6% and ORB by about 2.5 times.

KAZE and AKAZE are part of a group of methods with high values, though lower than SURF. In most sequences, they are characterized by either only slight changes or a moderate decrease in SCS as the number of keypoints increases. SIFT usually occupies an intermediate position between this group and the methods with low structural consistency. In most cases, the lowest metric values are observed for ORB, BRISK, and SuperPoint, with ORB often being the weakest method.

Thus, the SCS metric effectively reflects the correspondence between the spatial distribution of inliers and the scene's structure. While SuperPoint showed the best results in terms of CUI and RI, SURF is the clear leader in terms of SCS. This confirms that distinct spatial-structural metrics characterize different aspects of inlier quality and should therefore be considered jointly. At the same time, the absolute level of the metric also depends on the particular sequence, so its values should primarily be interpreted within each scene.

Fig. 8 presents the dependence of the penalized integral quality indicator Q_{pen} on the number of keypoints. Unlike the individual geometric and spatial-structural metrics, this

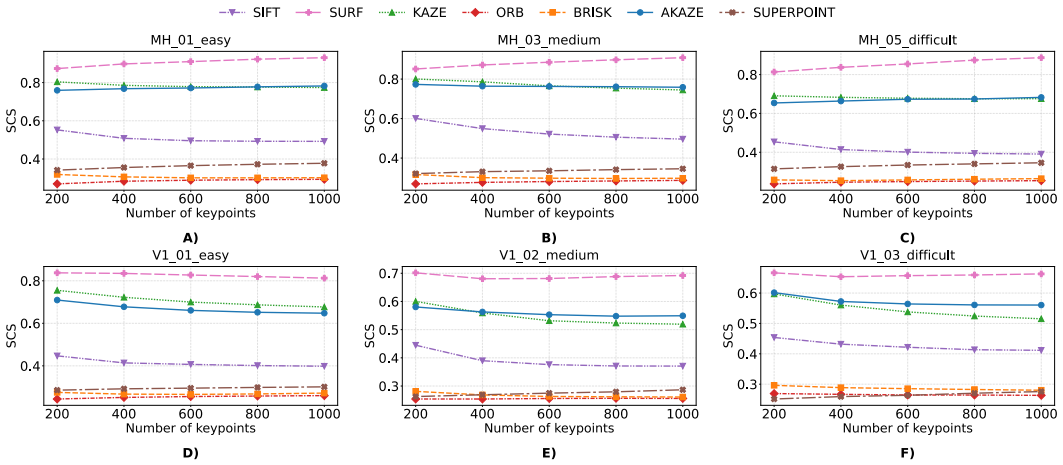

Fig. 7. Median values of SCS vs. the number of keypoints for different detectors in the EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

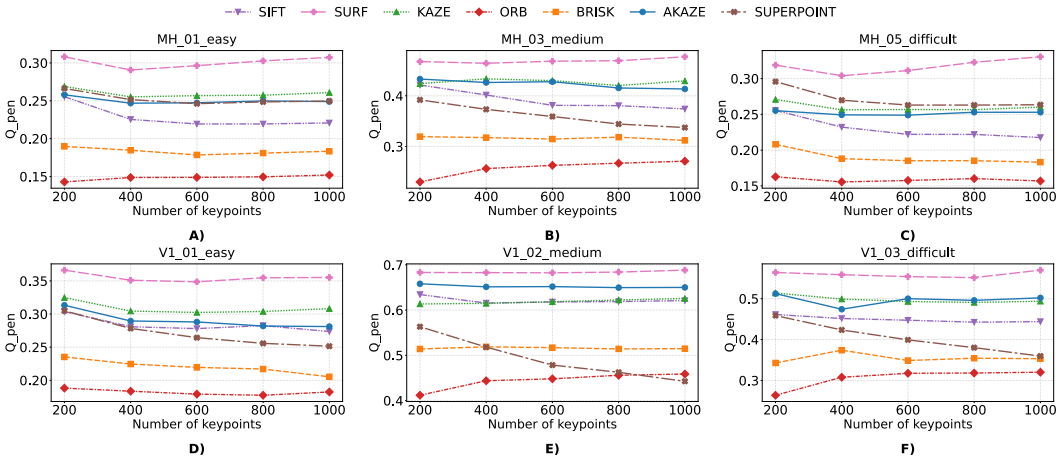

Fig. 8. Median values of Q_{pen} vs. the number of keypoints for different detectors in the EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

indicator summarizes their combined contribution and therefore provides a more comprehensive view of detector suitability in the monocular visual odometry task.

In most sequences, SURF consistently achieves the highest Q_{pen} values, indicating the best balance among geometric quality, spatial coverage, Q_{pen} consistency, and penalty components. For example, at N=1000, the Q_{pen} value for SURF is approximately 0.308 in MH_01_easy, 0.355 in V1_01_easy, and 0.567 in V1_03_difficult.

A second group of methods is usually composed of AKAZE and KAZE, which, in most cases, produce similar Q_{pen} values. For example, in V1_02_medium at N=1000, AKAZE reaches approximately 0.649, while KAZE and SIFT yield close but slightly lower values. A similar pattern is observed in V1_03_difficult, where AKAZE and KAZE also form the upper group after SURF, consistent with their high geometric metric values.

SuperPoint demonstrates non-uniform behavior. In some sequences, its Q_{pen} values are relatively high at small N, but then tend to decrease or stagnate. For example, in V1_03_difficult, the value for SuperPoint decreases from 0.459 at N=200 to 0.360 at N=1000, that is, by approximately 21.6%. This is in good agreement with the previously observed decrease in G_{ratio} for this method as the number of points increases.

ORB demonstrates the lowest Q_{pen} values in almost all sequences, while BRISK usually occupies an intermediate position between ORB and the group of stronger methods. For example, in MH_01_easy and V1_01_easy at N=1000, SURF exceeds ORB by approximately twofold. This confirms that ORB's weaker performance is observed not only in individual components but also in the overall assessment.

Thus, the Q_{pen} metric consistently summarizes the previous observations: SURF proved to be the most stable leader in terms of overall quality, AKAZE and KAZE formed a strong second group, whereas ORB and, to some extent, BRISK demonstrated lower overall suitability. SuperPoint, despite strong performance on some individual spatial metrics, remains inferior to the leaders due to its weaker geometric component, which becomes especially evident at large N. At the same time, the absolute level of the integral indicator also depends on the particular sequence, so its values should primarily be interpreted within each scene.

To analyze the relationship between the spatial-geometric indicators and odometry errors, Spearman rank correlation matrices were constructed separately for each EuRoC sequence, as shown in **Fig. 9**. This approach made it possible to reveal that both the strength and even the sign of the correlations may vary depending on the particular scene,

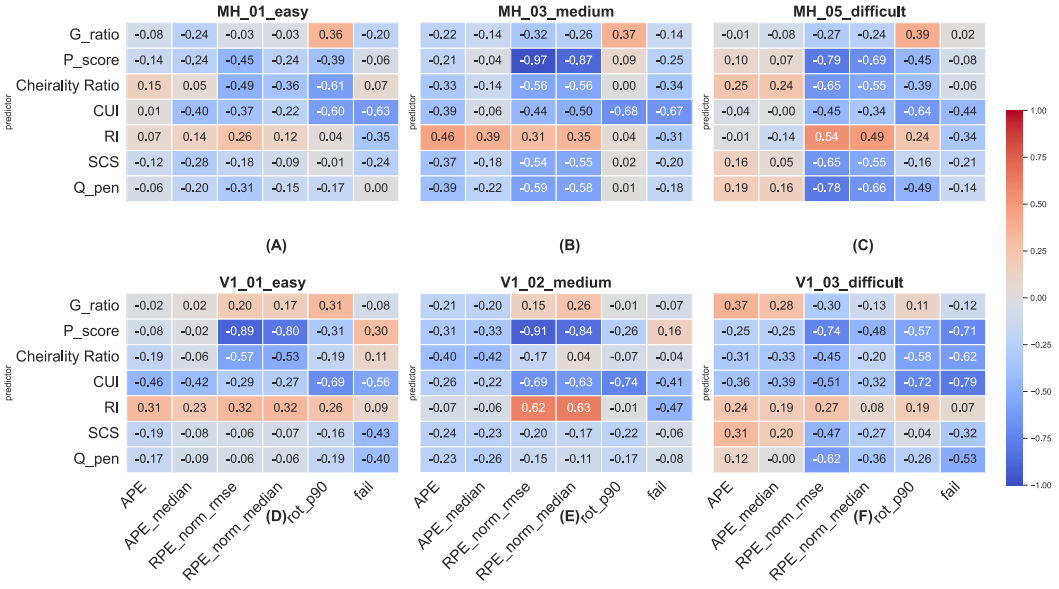

Fig. 9. Spearman rank correlation matrices between the spatial-geometric predictors and odometry error metrics for individual EuRoC sequences: (A) MH_01_easy, (B) MH_03_medium, (C) MH_05_difficult, (D) V1_01_easy, (E) V1_02_medium, (F) V1_03_difficult.

motion pattern, and observation conditions; that is, the informativeness of the metrics has a pronounced sequence-dependent character.

For the Machine Hall sequences, the most consistent relationship with odometry quality is demonstrated by the inlier coverage uniformity indicator RI, for which negative correlations are predominantly observed with translational errors, rotational tail-risk indicators, and failure rate. For example, for MH_01_easy, the correlation between the inlier coverage uniformity indicator and the failure rate is $\rho = -0.63$, and for MH_03_medium, it is $\rho = -0.67$. The local redundancy index RI, in contrast, shows mostly positive correlations with the error measures, consistent with the negative effect of local inlier clustering. For some sequences in this group, a pronounced relationship is also observed between the parallax indicator P_{score} and the normalized relative translational error RPE_{norm}; in particular, for MH_03_medium, the correlation is $\rho = -0.97$.

For the Vicon Room sequences, the strongest and most stable relationship with translational errors is most often observed with the parallax indicator, underscoring the importance of motion observability under these conditions. For example, for V1_02_medium, the correlation between the parallax indicator and the normalized relative translational error is $\rho = -0.91$, and for V1_03_difficult, it is $\rho = -0.74$. At the same time, the inlier coverage uniformity indicator remains substantially related to odometry stability; in particular, for V1_03_difficult, its correlation with the failure rate is $\rho = -0.79$. Other metrics, including the correct cheirality ratio, the inlier structural consistency indicator SCS, and the penalized integral quality indicator Q_{pen}, exhibit a less homogeneous pattern, indicating stronger dependence on the specific scenario.

Thus, the matrices for the individual sequences confirm that no single partial metric serves as a universal predictor of trajectory error. The most informative indicators for explaining changes in odometry quality were the inlier coverage uniformity indicator, the parallax indicator, and the normalized local redundancy indicator. The correlations themselves should be interpreted as indicators of a monotonic relationship rather than as direct evidence of causality.

The obtained results showed that different groups of metrics emphasize different aspects of the suitability of local features for the monocular visual odometry task. This is

especially evident in cases where a method exhibits strong spatial characteristics but lacks the same level of geometric reliability. In particular, SuperPoint proved to be one of the best methods in terms of coverage, uniformity and local redundancy. Yet its advantage was not preserved in the geometric metrics and the penalized integral indicator. This result indicates that good spatial coverage alone does not guarantee the best suitability for motion estimation.

In contrast, SURF was not always the unconditional leader across all individual partial metrics, yet it demonstrated the most stable balance among geometric consistency, structural correspondence, and integral assessment. This provides grounds for considering the balance of characteristics to be a more important property for practical use in VO than the maximization of any single criterion. In this sense, the results confirm that evaluating local features in visual odometry tasks should be based not on a single “best” indicator but on a set of complementary criteria.

At the same time, the correlation analysis showed that the informativeness of the partial metrics regarding odometry quality varies with the sequence type. In some scenes, uniform inlier coverage is more important, whereas in others, motion-observability characteristics are more informative. This means that there is no simple universal relationship between local feature quality and the final trajectory error. Such heterogeneity further confirms the relevance of a comprehensive evaluation approach that considers different metrics jointly. At the same time, the integral indicator is used as a means of generalized ranking rather than as a direct replacement for trajectory-based metrics.

CONCLUSION

This study presents a comprehensive investigation of local feature quality in monocular visual odometry using a combination of geometric, spatial-structural, and integral indicators. The performed analysis showed that no single metric is sufficient for a complete characterization of detector suitability for the VO task, since different indicators reflect different aspects of correspondence quality. According to geometric metrics, the best results were achieved by AKAZE, SURF, and partially by SIFT. In contrast, in terms of spatial-structural indicators, SuperPoint showed an advantage in inlier coverage uniformity and local redundancy, while SURF was superior in structural consistency.

It was shown that, as the number of keypoints increases, the results for most methods initially improve, but then often reach a saturation regime and in some cases even deteriorate. This indicates that the practical effectiveness of a detector is determined not only by the number of detected features, but also by their informativeness, spatial distribution, and ability to form geometrically valid correspondences. It was established that SURF is the most balanced method in the conducted experiments, as it consistently ranks among the leaders across different partial criteria and attains the highest values of the penalized integral quality indicator. AKAZE and KAZE formed a strong second group, whereas ORB showed the weakest results in most cases, both for the partial metrics and for the integral assessment.

The correlation analysis showed that the relationship between the partial metrics and the odometry quality indicators varies across sequences. For the Machine Hall group, the most informative indicator was the uniformity of inlier coverage. In contrast, for the Vicon Room group, the strongest relationship with translational errors was demonstrated by the parallax indicator. This confirms that no single metric can be regarded as a universal indicator of VO quality for all scene types, and that the integral approach is an appropriate means of multicriteria generalization and ranking.

Thus, the proposed approach to the spatial-geometric evaluation of local features enables not only comparison of methods but also explanation of their respective strengths and weaknesses in the context of visual odometry. The practical value of the obtained results lies in the ability to make a well-grounded detector choice based on the requirements of a particular application and the observation conditions.

Prospects for further research are associated with the use of other types of datasets, in particular those with more pronounced variations in illumination, texture, scene dynamics, and motion scale. It is also advisable to further investigate the behavior of the proposed indicators in combination with modern neural network-based detectors and descriptors, and to verify their suitability not only for monocular visual odometry but also for a broader range of tasks, including SLAM. A separate direction for future work will be to improve the integral indicator by adaptively adjusting the weighting coefficients based on the scene type.

ACKNOWLEDGMENTS AND FUNDING SOURCES

The authors received no financial support for the research, writing, and/or publication of this article.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any.

AUTHOR CONTRIBUTIONS

Conceptualization, [A.F., Yu.F.]; methodology, [A.F., Yu.F.]; validation, [A.F., Yu.F.]; writing – original draft preparation, [A.F.]; writing – review and editing, [A.F., Yu.F.]; supervision, [Yu.F.].

All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] Herrera-Granda, E. P., Torres-Cantero, A., Haro, G., & Nieto, M. (2024). Monocular visual SLAM, visual odometry, and structure from motion methods applied to 3D reconstruction: A comprehensive survey. Heliyon, 10(18), e37356. https://doi.org/10.1016/j.heliyon.2024.e37356

[2] Zhao, L., Li, Y., Wang, M., & Zhang, Y. (2025). PLL-VO: An efficient and robust visual odometry integrating point-line features and neural networks. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-G-2025, 1045-1052. https://doi.org/10.5194/isprs-annals-X-G-2025-1045-2025

[3] Luan, S., Yang, C., Qin, X., Chen, D., & Sui, W. (2024). Towards robust visual odometry by motion blur recovery. Frontiers in Signal Processing, 4, 1417363. https://doi.org/10.3389/frsip.2024.1417363

[4] Budzan, S., Wyżgolik, R., & Lysko, M. (2025). Performance analysis of keypoints detection and description algorithms for stereo vision based odometry. *Sensors, 25*(19), 6129. https://doi.org/10.3390/s25196129

[5] Huang, Q., Guo, X., Wang, Y., Sun, H., & Yang, L. (2024). A survey of feature matching methods. *IET Image Processing, 18*(6), 1385-1410. https://doi.org/10.1049/ipr2.13032

[6] Fesiuk, A., & Furgala, Y. (2025). Comprehensive spatial-geometric evaluation of keypoint detectors. *Electronics and Information Technologies, 32*, 67-86. https://doi.org/10.30970/eli.32.5

[7] Decker, P., Paulus, D., & Feldmann, T. (2008). *Dealing with degeneracy in essential matrix estimation*. In *2008 15th IEEE International Conference on Image Processing* (pp. 1964-1967). IEEE. https://doi.org/10.1109/ICIP.2008.4712167

[8] Ma, J., Jiang, X., Jiang, J., Zhao, J., & Guo, X. (2021). Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision, 129*(1), 23-79. https://doi.org/10.1007/s11263-020-01359-2

[9] Xu, S., Chen, S., Xu, R., Wang, C., Lu, P., & Guo, L. (2024). Local feature matching using deep learning: A survey. *Information Fusion, 107*, 102344. https://doi.org/10.1016/j.inffus.2024.102344
[10] Nagy, A., Barsi, Á., & Takács, B. (2025). A comparative evaluation of classical and deep learning visual odometry configurations. Engineering Proceedings, 113(1), 16. https://doi.org/10.3390/engproc2025113016
[11] Yu, J., Zhang, H., Liu, Q., & Chen, Z. (2025). Dynamic feature elimination-based visual–inertial odometry based on an optimized SuperPoint feature extractor. *Sensors, 26*(1), 52. https://doi.org/10.3390/s26010052
[12] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94
[13] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding, 110*(3), 346–359. https://doi.org/10.1016/j.cviu.2007.09.014
[14] Alcantarilla, P. F., Bartoli, A., & Davison, A. J. (2012). KAZE features. In *ECCV 2012* (LNCS 7577, pp. 214–227). https://doi.org/10.1007/978-3-642-33783-3_16
[15] Alcantarilla, P. F., Nuevo, J., & Bartoli, A. (2013). Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *BMVC 2013* (pp. 1–11). https://doi.org/10.5244/C.27.13
[16] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *ICCV 2011* (pp. 2564–2571). https://doi.org/10.1109/ICCV.2011.6126544
[17] Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). BRISK: Binary robust invariant scalable keypoints. In *ICCV 2011* (pp. 2548–2555). https://doi.org/10.1109/ICCV.2011.6126542
[18] DeTone, D., Malisiewicz, T., & Rabinovich, A. (2018). *SuperPoint: Self-supervised interest point detection and description*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 224-236). https://doi.org/10.48550/arXiv.1712.07629
[19] Budzan, S., Wyżgolik, R., & Lysko, M. (2025). Performance analysis of keypoints detection and description algorithms for stereo vision based odometry. *Sensors, 25*(19), 6129. https://doi.org/10.3390/s25196129
[20] Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M. W., & Siegwart, R. (2016). *The EuRoC micro aerial vehicle datasets*. *The International Journal of Robotics Research, 35*(10), 1157-1163. https://doi.org/10.1177/0278364915620033
[21] Lindenberger, P., Sarlin, P.-E., & Pollefeys, M. (2023). *LightGlue: Local Feature Matching at Light Speed*. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 17681-17692). https://doi.org/10.1109/ICCV51070.2023.01616
[22] Nannen, V., de Rezende, P. J., & Oliveira, M. M. (2013). *Grid-based spatial keypoint selection for real time visual odometry*. In *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP)* (pp. 1-8). https://doi.org/10.5220/0004270005860589
[23] Cho, H. M., Lee, S. H., & Kim, H. S. (2025). *Robust visual–inertial odometry via multi-scale deep feature extraction and adaptive keypoint selection*. *Applied Sciences, 15*(20), 10935. https://doi.org/10.3390/app152010935
[24] Howse, Joseph, and Joe Minichino. “Learning OpenCV 4 Computer Vision with Python 3: Get to grips with tools, techniques, and algorithms for computer vision and machine learning.”, *Packt Publishing Ltd*, 2020.

[25] Fesiuk, A., & Furgala, Y. (2025). Keypoint matches filtering in computer vision: Comparative analysis of RANSAC and USAC variants. *International Journal of Computing, 24*(2), 343–350. https://doi.org/10.47839/ijc.24.2.4018
[26] M. Ivashechkin, D. Baráth, J. Matas, "USACv20: Robust Essential, Fundamental and Homography Matrix Estimation," 2021. https://doi.org/10.48550/arXiv.2104.05044
[27] Nistér, D. (2004). *An efficient solution to the five-point relative pose problem*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26*(6), 756-770. https://doi.org/10.1109/TPAMI.2004.17
[28] Grupp, M. (2026). *evo: Python package for the evaluation of odometry and SLAM*. GitHub repository. https://github.com/MichaelGrupp/evo
[29] Umeyama, S. (1991). *Least-squares estimation of transformation parameters between two point patterns*. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13*(4), 376-380. https://doi.org/10.1109/34.88573
[30] Hartley, R., & Zisserman, A. (2004). *Multiple View Geometry in Computer Vision* (2nd ed.). Cambridge University Press.
[31] Hossein-Nejad, Z., & Nasri, M. (2017). *RKEM: Redundant keypoint elimination method in image registration*. *IET Image Processing, 11*(4), 273-284. https://doi.org/10.1049/iet-ipr.2016.0440
[32] Bailo, O., Rameau, F., Joo, K., Park, J., Bogdan, O., & Kweon, I. S. (2018). *Efficient adaptive non-maximal suppression algorithms for homogeneous spatial keypoint distribution*. *Pattern Recognition Letters, 106*, 53-60. https://doi.org/10.1016/j.patrec.2018.02.020
[33] Gauglitz, S., Foschini, L., Turk, M., & Höllerer, T. (2011). *Efficiently selecting spatially distributed keypoints for visual tracking*. In *2011 18th IEEE International Conference on Image Processing* (pp. 1869-1872). IEEE. https://doi.org/10.1109/ICIP.2011.6115832

ПРОСТОРОВО-ГЕОМЕТРИЧНЕ ОЦІНЮВАННЯ ЛОКАЛЬНИХ ОЗНАК У МОНОКУЛЯРНІЙ ВІЗУАЛЬНІЙ ОДОМЕТРІЇ

Андрій Фесюк, **Юрій Фургала***
Кафедра оптоелектроніки та інформаційних технологій
Львівський національний університет імені Івана Франка,
вул. Драгоманова 50, 79005 Львів, Україна

АНОТАЦІЯ

Вступ. Монокулярна візуальна одометрія (ВО) є важливим компонентом систем візуальної навігації, однак її точність залежить від якості локальних ознак і міжкадрових відповідностей. У задачі ВО важливими є не лише геометрична узгодженість, а й спостережуваність руху, фізична коректність відновленої конфігурації та просторово-структурні властивості локальних ознак. Метою роботи є комплексне оцінювання методів виявлення та опису особливих точок у монокулярній візуальній одометрії.

Матеріали та методи. Дослідження проведено на наборі даних EuRoC MAV. Проаналізовано методи ORB, BRISK, AKAZE, KAZE, SIFT, SURF та SuperPoint за кількості ключових точок від 200 до 1000. Оцінювання руху виконувалося на основі істотної матриці з використанням фільтра USAC_FAST, методу recoverPose, перевірки мінімального паралаксу та просторово керованого відбору ключових точок. Точність

відновленої траєкторії оцінювали за метриками APE та RPE. Для аналізу якості локальних ознак і відповідностей використовували геометричну складову, показник паралаксу, частку коректної хіральності, а також метрики рівномірності покриття особливими точками, локальної надлишковості та структурної узгодженості. Для узагальнення результатів застосовано інтегральний показник якості.

Результати. Геометричні метрики найчастіше виділяють алгоритми AKAZE та SURF, тоді як за просторовими характеристиками сильні позиції має метод SuperPoint. За показником структурної узгодженості відповідностей найкращі результати стабільно демонструє алгоритм SURF. Зі збільшенням кількості ключових точок для більшості методів спостерігається початкове покращення результатів із подальшим насиченням, а в деяких випадках – погіршення окремих характеристик. Найбільш збалансованим методом за сукупністю критеріїв виявився SURF, тоді як алгоритм ORB у більшості випадків продемонстрував найслабші результати. Кореляційний аналіз показав, що інформативність метрик залежить від типу послідовності.

Висновки. Запропонований підхід підтвердив доцільність багатокритеріального оцінювання локальних ознак у монокулярній візуальній одометрії. Показано, що жодна окрема метрика не є універсальною для всіх типів сцен, тоді як інтегральний показник дозволяє узагальнити різні аспекти якості та виконувати більш обґрунтоване ранжування методів.

Ключові слова: монокулярна візуальна одометрія, виявлення ключових точок, зіставлення зображень, оцінювання руху, глибоке навчання, нейронні мережі.

Received / Одержано	Revised / Доопрацьовано	Accepted / Прийнято	Published / Опубліковано
09 March, 2026	21 March, 2026	23 March, 2026	30 March, 2026