Journal of Intelligent & Robotic Systems (2024) 110:42
https://doi.org/10.1007/s10846-024-02077-4

REGULAR PAPER

How Challenging is a Challenge? CEMS: a Challenge Evaluation Module for SLAM Visual Perception

Xuhui Zhao[1] · Zhi Gao[1] · Hao Li[1] · Hong Ji[1] · Hong Yang[2] · Chenyang Li[1] · Hao Fang[3] · Ben M. Chen[4]

Received: 18 June 2023 / Accepted: 22 February 2024

Abstract
Despite promising SLAM research in both vision and robotics communities, which fundamentally sustains the autonomy of intelligent unmanned systems, visual challenges still threaten its robust operation severely. Existing SLAM methods usually focus on specific challenges and solve the problem with sophisticated enhancement or multi-modal fusion. However, they are basically limited to particular scenes with a non-quantitative understanding and awareness of challenges, resulting in a significant performance decline with poor generalization and(or) redundant computation with inflexible mechanisms. To push the frontier of visual SLAM, we propose a fully computational reliable evaluation module called CEMS (Challenge Evaluation Module for SLAM) for general visual perception based on a clear definition and systematic analysis. It decomposes various challenges into several common aspects and evaluates degradation with corresponding indicators. Extensive experiments demonstrate our feasibility and outperformance. The proposed module has a high consistency of 88.298% compared with annotation ground truth, and a strong correlation of 0.879 compared with SLAM tracking performance. Moreover, we show the prototype SLAM based on CEMS with better performance and the first comprehensive CET (Challenge Evaluation Table) for common SLAM datasets (EuRoC, KITTI, etc.) with objective and fair evaluations of various challenges. We make it available online to benefit the community on our website.

Keywords Robotics · Resilient SLAM · Visual degradation · Challenge evaluation

1 Introduction

With the booming development of robotics in our AI era, more intelligent unmanned systems are deployed to numerous scenes and play non-substitutable roles in various applications, including space exploration with Mars rovers, air reconnaissance with drone swarms, ground service with smart vehicles, subterranean rescue with quadrupedal robots, and underwater archaeology with submersibles. Generally, SLAM is the core component in an intelligent unmanned system and sustains autonomy fundamentally, where its robustness to visual challenges determines the performance of the whole system [1]. From the first to the state-of-the-art SLAM, most achieve inspiring performance, essentially relying on the high-quality perception of the ambient environment. Unfortunately, the perception usually degrades dramatically in hard or inaccessible scenes where we deploy intelligent unmanned systems [2–4]. Meanwhile, we typically cannot predict what visual challenge will come next during exploration in unknown environments.

Many efforts are devoted to this topic, which mainly involves the SLAM field and IQA/VQA (Image/Video Quality Assessment) field. In the visual SLAM field, researchers have focused on robust performance in challenging environments and generally propose condition-blind and condition-aware methods. The condition-blind methods typically enhance certain visual qualities (such as illumination) merely and constantly for stable tracking regardless of the ambient environment change, resulting in limited robustness in a changing world and redundant computation burden for load platforms. While the condition-aware methods are usually capable of specific visual degradation with qualitative and straightforward metrics, leading to a satisfying SLAM performance in simple and toy applications. However, despite

✉ Zhi Gao
gaozhinus@gmail.com

Extended author information available on the last page of the article

Published online: 09 March 2024

the good idea, current simple metrics are in an early stage without clear definitions and typically only for one visual challenge, making it infeasible for various challenges in a complex world. Therefore, there are few robust SLAM methods for practical challenging environments due to the lack of general and systematic evaluation of visual degradation. Besides the SLAM field, IQA/VQA methods typically evaluate images from aspects of human visual perception rather than machine vision, where visually pleasing images are not necessarily beneficial to visual SLAM algorithms. Moreover, these methods usually focus more on accuracy and take few considerations for real-time performance that is of great importance for practical online tasks, resulting in less feasibility for SLAM. Therefore, in spite of various works, the full autonomy of unmanned systems in challenging scenes remains an open problem without general and elegant solutions. That is why CVPR 2020 held the SLAM Challenge [5] to break the limits of visual SLAM in various challenging scenarios.

Generally, the word "challenging" is subjective and ambiguous, and the research on resilient visual SLAM is also a complex task. Nevertheless, it is a topic that must be further studied since we are convinced that the key to the aforementioned problem is the quantitative awareness of various challenges. In practical applications, the ambient environment is usually unknown, especially in the search and rescue tasks. The SLAM system has to face any sudden visual challenges robustly with effective algorithms to keep continuous tracking. Otherwise, the system either crashes or has to conduct much redundant computation constantly to stay alive. For example, the SLAM cannot run in a low-texture scene with illumination enhancement methods, where many current SLAM methods are in this dilemma due to the lack of effective challenge evaluation. We think an ideal robust SLAM towards adverse environments should be capable of "targeted and on-demand enhancement" and answer these two questions: (1) Do we need visual enhancement? (2) Which enhancement should we conduct? For example, we conduct illumination enhancement in low-light frames and switch to feature enhancement if low-texture frames occur, otherwise, no enhancement is conducted. In this way, we can save unnecessary computation in normal scenes while improving robustness in challenging scenes, balancing accuracy, efficiency, and robustness. To achieve this goal, the quantitative and accurate perception of various visual degradation plays a crucial role in SLAM. However, existing work in both SLAM and IQA/VQA fields generally still has much room to improve. Therefore, we propose the CEMS (Challenge Evaluation Module for SLAM) to push the boundary. We first define "visual challenges" from the robustness and accuracy of SLAM. Then we derive a framework from the classic imaging process to decompose challenges into several common aspects for a general analysis inspired by chromatic dispersion (Fig. 1). Consequently, we develop the CEMS and evaluate various challenges in SLAM datasets. Finally, based on the proposed CEMS, we demonstrate a prototype SLAM robust to adverse illumination with better efficiency. Moreover, we conduct qualitative challenge evaluation on over 1.45 million frames from public SLAM datasets and obtain insightful tips with CET (Challenge Evaluation Table). To our best knowledge, few similar works exist, and our contributions are below:

- A clear definition and systematic analysis of various challenges are proposed for visual SLAM from the classic imaging process in computer vision.

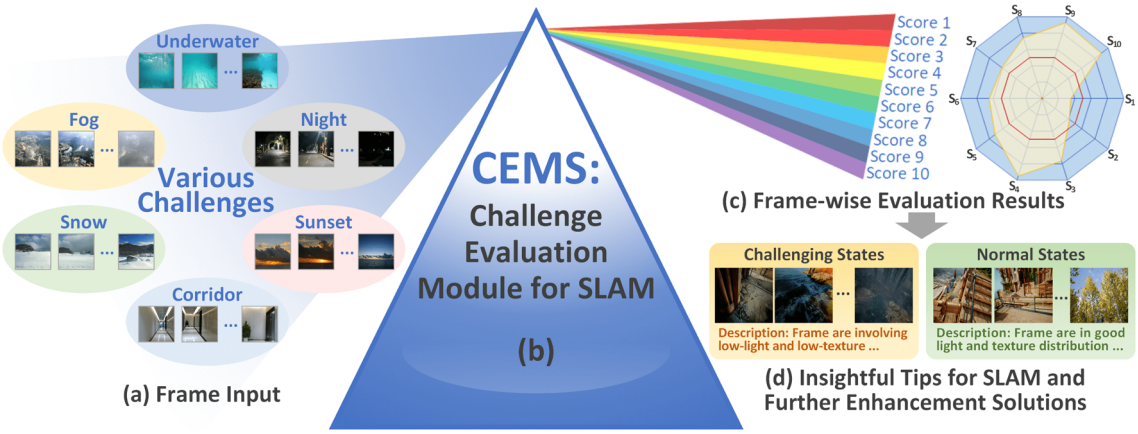

Fig. 1 The intuition of our proposed (b) Challenge Evaluation Module for SLAM (CEMS). Despite (a) various challenges in input frames, we decompose them into several common aspects in the classic imaging process with the inspiration from chromatic dispersion and calculate (c) perception scores correspondingly, providing (d) evaluation results with insights for SLAM. The CEMS enables SLAM to intelligently react to different changes with better robustness and efficiency in practical applications, following the "targeted and on-demand enhancement" idea

- A general evaluation module for visual SLAM (CEMS) is developed to objectively and quantitatively recognize visual challenges in practical applications.
- Initial SLAM with better performance and challenge evaluation table (CET) with insights suggest the potential of CEMS for online and offline tasks, respectively.

2 Related Work

Following the same taxonomy in Section 1, we categorize related methods into the SLAM field (Section 2.1) and IQA/VQA field (Section 2.2) and elaborate on them respectively.

2.1 Resilient SLAM Towards Visual Challenges

We divide existing SLAM proposed for visual challenges into condition-blind and condition-aware methods according to the evaluation of visual degradation.

2.1.1 Condition-blind Methods

These methods believe that better visual perception brings benefits to SLAM [6] and enhances the quality of frames, making the sequences less "challenging", thus improving the performance of visual SLAM. However, many do not evaluate the challenges explicitly. For example, the SINV2 augments the SLAM pipeline for underwater environments with histogram equalization [7]. For low-texture scenes, EDPLVO combines the point and line features with photometric error [8], while the MEGVII team uses SuperPoint [9] and SuperGlue [10] for robust feature matching, winning the first prize in CVPR 2020 SLAM Challenge. Even latent 3D information and structures in man-made environments are explored with Manhattan and Atlanta world assumptions for robust pose estimation [11, 12]. Besides, to enhance the data association, AirDOS improves the performance in dynamic scenes with articulated objects [13], and a neural network is trained to control imaging parameters to eliminate lighting changes [14]. Despite promising results of these methods in targeted scenes, they usually lead to unintelligent solutions and non-optimal performance in practical and complex applications with the following two significant drawbacks. First, constant enhancement regardless of external conditions usually results in unnecessary and redundant computation which is typically not affordable, especially for resource-limited platforms. For example, we obviously do not have to enhance perception in good conditions in most cases. Second, these methods usually have excessive focus on specific challenges and are not capable of tackling different challenges, making the SLAM fragile in scenes out of design scope. For example, the SLAM with illumination enhancement may crash in low-texture environments. Therefore, there is still a gap between these methods and robust applications in challenging environments due to the lack of automatic degradation evaluation.

2.1.2 Condition-aware Methods

Compared with condition-blind methods, these methods usually perceive challenges with indirect or qualitative evaluation. We further divide them into frame-based and pose-based methods, where the evaluation is conducted on input frames before tracking and estimated poses after tracking, respectively.

Frame-based Methods These methods usually focus on specific scenes and evaluate the quality of input frames directly with straightforward metrics. For example, to discriminate the airborne dust and smoke for the outdoor perception of UGVs (Unmanned Ground Vehicles), a method leveraging on Shannon information measurement is proposed in [15, 16], then a prototype visual SLAM is developed for these adverse conditions [17–19]. Similarly in [20], an illumination change recognition method is proposed for the robust visual localization of Astrobee in ISS (International Space Station) day and night. Generally, these methods are more beneficial to SLAM compared with pose-based ones since we can get detailed degradation information and adequate time for corresponding enhancement before pose estimation. While we usually cannot derive the specific reason only from estimated poses in pose-based methods since multiple factors can result in the same degradation in pose. Despite the good idea, the main problem of these methods is the simple challenge evaluation with limited focus on the various and changing degradation. Adopted metrics are usually not systematic and general for a complex world, resulting in limited performance improvement in practical scenes. For example, the SLAM with only illumination evaluation cannot perceive texture degradation thus the system has a high risk of crash in this condition.

Pose-based Methods Pose-based methods usually achieve degradation awareness by checking the health state of odometry with concessive pose estimations. For example, in the DARPA Subterranean (SubT) Challenge [21], the CERBERUS framework (proposed by ETH [22]) considers the D-optimality criterion as the metric to quantify the uncertainty of the robot and the covariance matrix of visual odometry estimates [23]. Similarly, the NeBula framework (proposed by NASA [24]) offers the HeRO [25] to expect and detect failures in perceptually degraded operating conditions, where the health check is mainly accomplished with odometry estimations [26]. Then the sensor fusion is achieved by LiDAR frontend LOCUS [27] and LION [28]. Finally, the factor-graph-based SLAM LAMP [29] performs global localization and mapping with a robust loop clo-

sure module called DARE-SLAM [30]. Moreover, team CTU-CRAS-NORLAB coordinates wheeled robots, tracked robots, crawling robots, and aerial robots for resilient exploration [31]. Generally, these methods are suitable for the health state estimation in odometry. The focus is mainly next moment-focused, where the health states are used for different sensor switching and fusion when the next input comes in the following tracking. However, we cannot obtain explicit reflection of current visual challenges since different degradations usually lead to the same decline in odometry, which is a many-to-one mapping. Therefore, many SLAM methods tackle challenging environments subjectively with inexplicit awareness of degradation, which is a deficiency for resilient visual localization towards various hard scenes.

2.2 IQA/VQA for Visual Quality

IQA/VQA methods also focus on the visual quality of images/videos, which is related to our topic to some extent. IQA can be categorized into full-reference (FR), reduced-reference (RR), and no-reference (NR) methods according to the dependence on information, where the full, partial, and no original image is accessible, respectively. Generally, NR methods try to evaluate images with the imitation of HVS (Human Vision System) and cover a wide range from classic FSIM [32] and natural scene statistics [33] to popular deep neural network [34], meta learning [35] and transformer [36]. Many methods are proposed for different purposes. For example, an IQA method for underwater scenes is proposed with frequency transformation in [37], and an evaluation method for night-time images is proposed from both subjective and objective aspects in [38]. Other works include fogging assessment [39], image noise [40], image compression [41] and multi-exposure fusion [42]. Despite devoted efforts and satisfying performance, they may not be suitable for challenge evaluation in SLAM due to few considerations on real-time performance and different definitions of "image quality", where visually pleasing scenes may be adverse to SLAM (such as the beautiful sunset over the sea as shown in Fig. 1(a)). Besides, we need to evaluate the temporal change between successive frames with real-time requirements for SLAM, while they usually focus on a single image. Although VQA methods consider temporal information, they usually quantify the compression quality [43] with high computational complexity. Therefore, IQA/VQA methods are not directly applicable to the challenge evaluation in visual SLAM, while some classic techniques and ideas may also be adopted.

As a brief summary of related work, the quantitative, systematic, and general evaluation of visual challenges for SLAM is still in a very early stage. Methods in the SLAM field are usually too simple in practical applications with limited robustness and methods in the IQA/VQA field are typically not suitable for SLAM with different focuses. Despite the difficulties and the lack of systematic works [28] in visual SLAM, it is still worth evaluating challenges quantitatively, which will bring significant improvements to SLAM performance in degraded environments, as demonstrated in LiDAR SLAM [44, 45]. Therefore, we propose the general evaluation module of visual challenges with qualitative analysis for resilient and robust SLAM in adverse scenes.

3 The Proposed Method

We first define the visual challenges for SLAM, propose the general analysis framework (Section 3.1), then present the quantitative evaluation module with ten selected scores (Section 3.2). Finally, we give implementation details for the module (Section 3.3).

3.1 General Definition and Analysis Framework for Challenges

What are the challenges for visual SLAM? It may be simple if your answer is the scene where SLAM fails from a posteriori view. However, the conclusions obtained by deducing the causes from the results usually provide little comprehensive and insightful information for practical applications. In fact, the general definition of challenging scenarios is a challenge indeed due to innumerable degradations and varying goals in various tasks. Here, we categorize challenges for intelligent unmanned systems into two types: challenges for exploration and challenges for perception, where the former has been studied with satisfying results [23, 46] in the control field but the latter is the opposite, leaving many open but valuable problems. To propose a clear definition, we start with two necessary hypotheses for the most available SLAM:

- H1: "Reliable association" - frame-wise matching should be performed with adequate, distinguishable, and consistent features for reliable correspondence.
- H2: "Static scene" - objects in scenes should be relatively static to each other with consistent relationships except the camera to ensure physical correctness of tracking.

Generally, vision is the only input for visual SLAM, thus robust visual matching guarantees accurate pose estimation and plays a crucial role in reliable tracking (H1 hypothesis). Many external factors may influence the H1, such as adverse illumination, low textures, and scene changes over a large timeframe. Despite many enhancement methods being proposed, it is still expected to have better visual perception inputs for tracking. Moreover, the classic visual

SLAM framework typically cannot handle dynamic objects and may output trajectories that do not conform to the facts. Since moving objects may break the latent assumption for static scenes due to relative motion (H2 hypothesis). Despite we can filter moving objects with various algorithms, the ultimate goal is to leverage the remaining static parts for robust tracking. This further implies the importance of the H2 hypothesis to SLAM. In practical applications with robots, various challenges usually break either or both of these two assumptions with the degradation of frame-wise association or movement ambiguity, resulting in the corresponding impacts on SLAM:

- Declined robustness for vision-based pose estimation with different anomalies (gap, divergence, and jump as defined in [25]) in challenging scenes.
- Poor accuracy and smoothness of estimated trajectories or trajectories that looks good but do not conform to the fact motion and ground truth.

Therefore, we propose the definition based on these hypotheses: *A visual perception is regarded as challenging if it breaks hypothesis H1 or H2 or both, decreasing the robustness and accuracy of the SLAM*. It is worth noting that we focus on natural and non-collaborative scenes rather than the ones with manually enhanced objects, such as ArUco and AprilTags artificial markers. Moreover, we focus on visual challenges in perception rather than challenges brought by the deficiency of algorithms. For example, some rare scenes are in good condition but still challenging to end-to-end SLAM with limited generalization. We further analyze various challenges based on this definition, and derive corresponding scores. It is generally tricky and infeasible to enumerate every challenging phenomenon in our complex world to formulate a general analysis framework. Inspired by chromatic dispersion, which indicates that light is composed of a finite spectrum with different wavelengths, we propose a framework to act as the special prism, decomposing various physical challenges to hierarchically common aspects and their combinations (Fig. 1). From the view of classic computer vision, the ambient perception in SLAM is essentially the well-studied imaging process [47]. Therefore, we first decompose various challenges that break the aforementioned hypotheses into illumination, scene, and sensor aspects, which comprise the classic imaging system. Then, we break down each aspect with several indicators and propose an analysis framework for visual challenges, as shown in Fig. 2.

3.2 Quantitative Challenge Evaluation Module

Following the general analysis framework, we derive ten scores in three aspects for challenges and design a quantitative evaluation module, where the inputs are only images, and the outputs are ten evaluation scores $S_i\,(i \in [1, 10] \cap \mathbf{N}^+)$, overall perception score S_p, and final judgments with tips. Generally, the higher S_p indicates better perception quality for SLAM with fewer challenges. To accurately fit the trend of various scores while keeping the computation burden at a low level, we adopt the three-segmented linear function as a paradigm for the general scoring, as defined in Eq. 1.

$$Score(x;\ \mathbf{C}) = \begin{cases} a_1 x + b_1 & t_1 \le x < t_2 \\ a_2 x + b_2 & t_2 \le x < t_3 \\ a_3 x + b_3 & t_3 \le x \le t_4 \end{cases}, \tag{1}$$

where $\mathbf{C} = \{a_1, a_2, a_3, b_1, b_2, b_3, t_1, t_2, t_3, t_4\}$ is the set of coefficients in the scoring function $Score(\cdot)$, a and b refer to coefficients for line segments, t indicates end points of each segment. $\mathbf{C_i}(i \in [1, 10] \cap \mathbf{N}^+)$ refers to corresponding coefficients for ten scores.

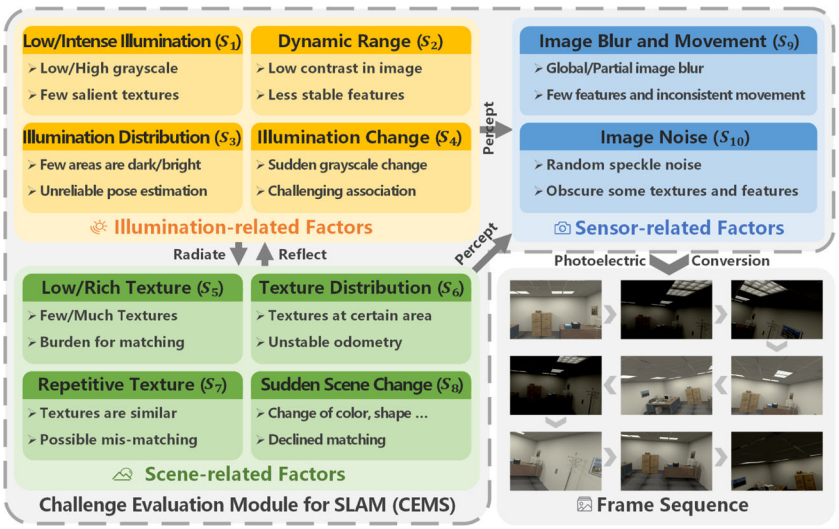

Fig. 2 The analysis framework and corresponding scores for visual challenge evaluation (Images are from ICL-NUIM dataset [48]). We break down various physical challenges into three factors in the classic imaging process, each with several scores for detailed quantitative evaluation

3.2.1 Illumination-related Scores

Good illumination is one of the prerequisites for a robust SLAM. From spatial and temporal perspectives, the low or intense illumination (S_1), the low dynamic range (S_2), the uneven illumination distribution (S_3), and the sudden frame-wise illumination change (S_4) typically bring challenges to feature extraction and data association in SLAM.

Low/Intense Illumination (S_1). Based on the gray world assumption [49], it is usually harmful to visual tracking with low or intense illumination. Therefore, we follow the proposed (1) and design the S_1 score curve as Eq. 2. For 8-bit images, the grayscale of 128 is considered the best illumination, while 0 and 255 are the opposite (too dark or too bright for an effective visual perception).

$$S_1 = Score(G_{mean};\ \mathbf{C_1}), \tag{2}$$

where $G_{mean} \in [0, 255]$ is the mean grayscale of the current frame $\mathcal{F}$ and calculated by pixel-wise operation. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_1}$ is the coefficient set for S_1.

Dynamic Range (S_2). Generally, an image with a small dynamic range usually has poor contrast, resulting in the challenge of extracting stable and distinguishable visual features. To avoid unexpected extreme grayscales, we truncate the central 96% of the whole histogram to calculate the valid dynamic range according to the Gaussian distribution. Finally, the S_2 score for 8-bit images is as Eq. 3.

$$S_2 = Score(G_{\max} - G_{\min};\ \mathbf{C_2}), \tag{3}$$

where $G_{\min}, G_{\max} \in [0, 255]$ are the lower and upper bounds of valid grayscale range in current frame $\mathcal{F}$. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_2}$ is the coefficient set for S_2.

Illumination Distribution (S_3). The evenly distributed illumination usually leads to a more uniform distribution of visual features, which is beneficial for SLAM [50]. We first divide the input image into grids and represent each grid with a Gaussian-weighted grayscale. Then, we count the number of grids with the same grayscale and calculate the corresponding variance for illumination distribution. Finally, the S_3 score for 8-bit images with 32×32 grids are calculated with segmented function as Eq. 4.

$$S_3 = Score(\frac{\sum_{j=0}^{255}(N_j - N_{mean})^2}{256};\ \mathbf{C_3}), \tag{4}$$

where $N_j \in [0, 1024]$, $j \in [0, 255]$ indicates the grid number at grayscale j and N_{mean} refers to the mean of all N_j. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_3}$ is the coefficient set for S_3.

Illumination Change (S_4). Significant illumination change may also lead to the failure of frame-wise data association. We evaluate this sudden change by comparing the absolute value of the mean grayscale difference in two consecutive frames, where the scoring function S_4 is as Eq. 5 for 8-bit images.

$$S_4 = Score(abs(G^P_{mean} - G_{mean});\ \mathbf{C_4}), \tag{5}$$

where $G^P_{mean}, G_{mean} \in [0, 255]$ is the mean grayscale of previous frame $\mathcal{F}^{\mathbf{P}}$ and current frame $\mathcal{F}$, respectively. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_4}$ is the coefficient set for S_4.

3.2.2 Scene-related Scores

The scene or photographed objects also play an essential role in the imaging process and significantly affect the performance of SLAM. From the view of frame-wise data association, the number of features (S_5), the distribution of features (S_6), the repetitive features (S_7), and the sudden scene change between frames (S_8) make the environment challenging for visual SLAM.

Low/Rich Textures (S_5). The proper sparsity of features benefits the SLAM rather than extreme states - low texture brings challenges to feature extraction, while rich texture leads to the decline of matching and efficiency. We take grayscale gradient as the essential metric. Specifically, we obtain a pixel-wise gradient map of current frame $\mathcal{F}$ with the classic Sobel operator. Then, we calculate the S_5 as Eq. 6 for texture evaluation according to the mean gradient in this map.

$$S_5 = Score(mean(\bigtriangledown^2\mathcal{F});\ \mathbf{C_5}), \tag{6}$$

where $mean(\cdot)$ indicates the averaging operation for the gradient map, $\bigtriangledown^2$ represents the Sobel operation. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_5}$ is the coefficient set for S_5.

Texture Distribution (S_6). Basically, it is easier to get accurate poses with even-distributed frame-wise association rather than local-distributed textures [45, 51]. We divide the gradient map obtained from S_5 into 16×16 grids and regard the grid as valid if its gradient sum exceeds the given threshold (we set 100 here). Finally, we calculate the S_6 score for distribution evaluation based on the proportion of valid grids N_{val} to all grids ($N_{val} + N_{inv}$), as written in Eq. 7.

$$S_6 = Score(\frac{N_{val}}{N_{val} + N_{inv}};\ \mathbf{C_6}), \tag{7}$$

where N_{val} and N_{inv} are the number of valid and invalid grids in current frame $\mathcal{F}$, respectively. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_6}$ is the coefficient set for S_6.

Repetitive Texture (S_7). Similar textures influence the performance of matching and thus have an impact on pose estimation. We quantitatively evaluate repetitive textures leveraging on GLCM (Grey Level Co-occurrence Matrix),

where the bigger homogeneity indicator I_{hom} in GLCM indicates more repetitive textures in an image. Then the S_7 score is calculated with Eq. 8.

$$S_7 = Score(I_{hom};\ \mathbf{C_7}), \tag{8}$$

where $I_{hom} \in [0, 1]$ is the homogeneity defined in [52]. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_7}$ is the coefficient set for S_7.

Sudden Scene Change (S_8). Due to the high frequency of vision cameras, the change between successive frames is usually tiny. However, sudden scene change may exist due to rapid motion, moving objects, and even long-term revisiting. Therefore, we quantify this sudden scene change by the similarity of corresponding histograms. Specifically, for 8-bit images, we calculate the S_8 score by iterative comparing the pixel number of each grayscale level j in channel $k \in \{Red(0), Green(1), Blue(2)\}$, as Eq. 9.

$$S_8 = Score(\sum_{k=0}^{2}\sum_{j=0}^{255} 1 - \frac{abs(N_{jk}^{P} - N_{jk})}{\max(N_{jk}^{P}, N_{jk})};\ \mathbf{C_8}), \tag{9}$$

where N_{jk}^{P} indicates the pixel number of grayscale j in channel k within the previous frame $\mathcal{F}^{\mathcal{P}}$, while the N_{jk} refers to the counterpart in the current frame $\mathcal{F}$. $abs(\cdot)$ and $\max(\cdot)$ denote the operation of obtaining absolute value and max value, respectively. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_8}$ is the coefficient set for S_8.

3.2.3 Sensor-related Scores

Last but not least, camera sensors convert photons into electrical signals in the imaging process and also play an essential role in SLAM. Generally, the image blur and dynamic objects (S_9) and the image noise (S_{10}) are harmful to stable feature extraction and matching in SLAM.

Image Blur and Movement (S_9). The image blur and inconsistent movement led by moving objects threaten the reliable feature matching in SLAM. We think clear images change a lot after intentionally blurring, while blurry images are resistant to it. Therefore, we blur the current input frame $\mathcal{F}$ with a Gaussian kernel and get an intentionally blurred image $\mathcal{F}_{\mathbf{B}}$. Then, we conduct the Sobel operation for gradient maps of both $\mathcal{F}$ and $\mathcal{F}_{\mathbf{B}}$. The image blurring is evaluated by the similarity indicator $I_{blur} = SSIM(\nabla^2\mathcal{F}, \nabla^2\mathcal{F}_{\mathbf{B}}) \in [0, 1]$ (Structure Similarity Index Measure) of these two gradient maps. Smaller I_{blur} indicates better perception with less image blurring. Besides blurring, moving objects also bring inconsistent movement between frames. We calculate the consistency for evaluation, in which the general transformation is calculated by feature matching with RANSAC. The the number of outliers N_{out} and inliers N_{in} is used to calculate the outlier proportion indicator $I_{move} = \frac{N_{out}}{N_{in}+N_{out}} \in [0, 1]$. Smaller I_{move} indicates more consistent movement. Finally, the S_9 score for both blurring and moving is calculated with I_{blur} and I_{move} as writtern in Eq. 10.

$$\begin{aligned} S_9 &= \frac{1}{2}(Score(I_{blur};\ \mathbf{C_{blur}}) + Score(I_{move};\ \mathbf{C_{move}})) \\ &= \frac{1}{2}(Score(SSIM(\nabla^2\mathcal{F}, \nabla^2\mathcal{F}_{\mathbf{B}});\ \mathbf{C_{blur}}) \\ &\quad + Score(\frac{N_{out}}{N_{in}+N_{out}};\ \mathbf{C_{move}})), \end{aligned} \tag{10}$$

where $SSIM(\cdot)$ is the similarity calculation, ∇^2 denotes the Sobel operation. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_9} = \{\mathbf{C_{blur}}, \mathbf{C_{move}}\}$ is the coefficient set for S_9.

Image Noise (S_{10}). The image noise usually hinders the robust description and matching of features in SLAM. Similar to S_9 score, we compare the difference of current frame $\mathcal{F}$ and its median filtered result $\mathcal{F}_{\mathbf{N}}$. Finally, the S_{10} score is calculated with the mean difference of $\mathcal{F}$ and $\mathcal{F}_{\mathbf{N}}$ as Eq. 11.

$$S_{10} = Score(mean(abs(\mathcal{F} - \mathcal{F}_{\mathbf{N}}));\ \mathbf{C_{10}}), \tag{11}$$

where $abs(\cdot)$ and $mean(\cdot)$ denote the operation of obtaining absolute value and mean value, respectively. $Score(\cdot)$ is defined in Eq. 1, $\mathbf{C_{10}}$ is the coefficient set for S_{10}.

3.2.4 The Overall Perception Score

The overall perception score S_p is calculated with all aforementioned ten scores as Eq. 12, which the lower indicates more challenging states.

$$S_p = \alpha\sum_{i=1}^{4} S_i + \beta\sum_{i=5}^{8} S_i + \gamma\sum_{i=9}^{10} S_i, \tag{12}$$

where α, β, and γ are weights for illumination, scene, and sensor aspect, respectively. Finally, we regard the scene a challenge if $\forall S_i < T_l \vee S_p < T_h$, where T_l and T_h are given thresholds for separate scores S_i and overall perception score S_p. Some visualization of different challenging level samples evaluated by CEMS is shown in Fig. 3. It can be seen that the automatically generated evaluation results are generally consistent with our experience. For example, for sudden illumination change (S_4), the evaluated bad result is with significant lighting difference as the yellow eclipse shows while the fair result has moderate change. The evaluated excellent result almost has no illumination changes which is beneficial for visual SLAM. More evaluation results can be found on our website https://gaozhinuswhu.com.

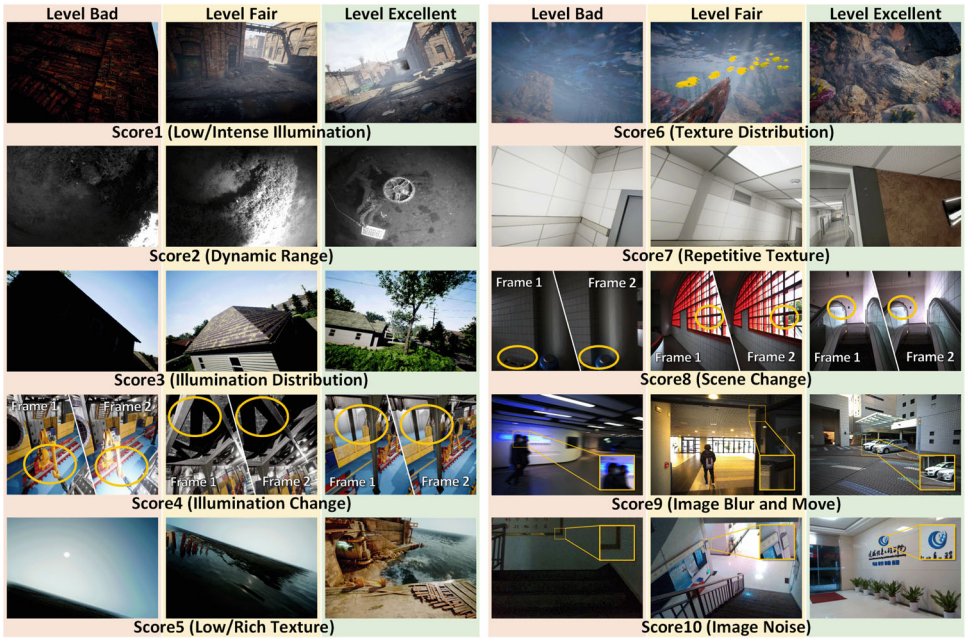

Fig. 3 Visualization of different score evaluation results automatically generated from CEMS. The results are briefly divided into bad (red), fair (yellow), and excellent (green) levels

3.3 Implementation Details

For accurate estimation of scoring functions and performance evaluation, we need to annotate diverse challenging samples based on the proposed framework manually and fit corresponding coefficients. The recommended process mainly involves three steps, as shown in Fig. 4. First, we define five perception levels (bad, poor, fair, good, and excellent) according to score ranges. Also, we select several SLAM experts for initial scoring. We adhere to the Recommendation ITU-R BT.500-14 [53] proposed by the ITU (International Telecommunication Union) and adopt SSCQE (Single Stimulus Continuous Quality Evaluation) for scoring, where the experts score the consecutive input frames within a duration according to their experience in visual SLAM and perception without any references. It should be noted that the scoring is achieved on the visual challenges for SLAM rather than aesthetic for human eyes as IQA/VQA does. Then, the raw annotations will be cleaned to filter significant errors and we calculate the MOS (Mean Opinion Score) with further rejection of outliers. Thus, the coefficients of different scoring functions can be fitted via iterative optimization according to given paradigms as defined in Eqs. 1, 2, 3, 4, 5, 6, 8, 9, 10 and 11. Finally, we output and save coefficients for each score (take S_1 as an example, as shown in Fig. 4), achieving the complete evaluation module for SLAM. More specifically, for general scoring function estimation $S_i (i = 1, 2, \cdots, 10)$, we hire three experts and select one

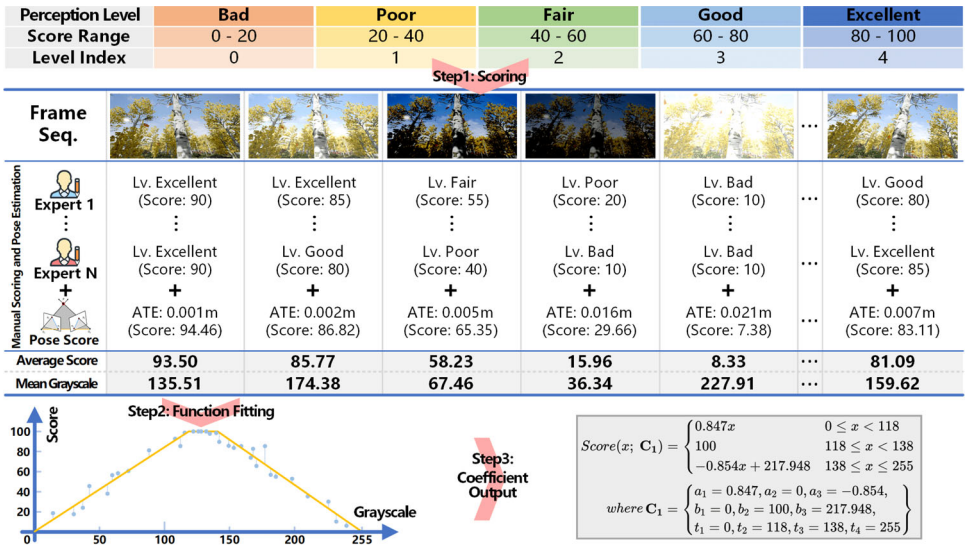

Fig. 4 Illustration of estimation for scoring function (take S_1 for example). We divide all states into five levels and obtain scores by manually scoring and pose estimation. The coefficient $\mathbf{C_1}$ is then fitted with minimal error

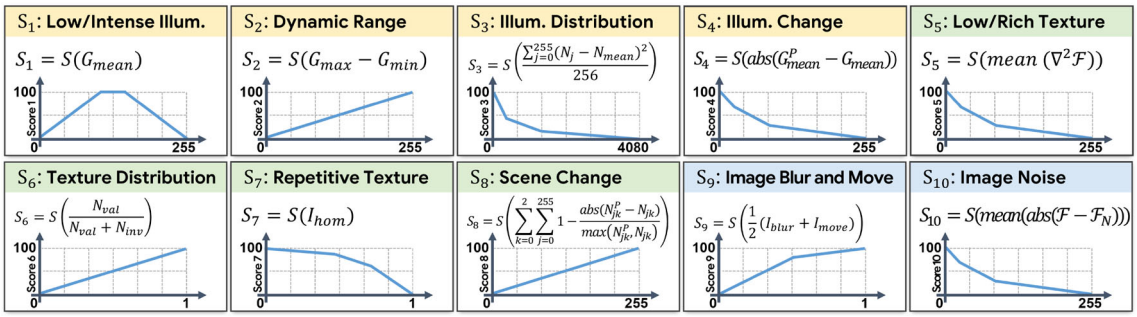

Fig. 5 Visualization of each scoring function in CEMS for quantitative challenge evaluation. Yellow rectangles indicate S_1 - S_4 in illumination aspect, green rectangles indicate S_5 - S_8 in scene aspect, and blue rectangles indicate S_9 - S_{10} in sensor aspect

thousand representative images in total from the public, simulated, synthesized, and field-collected sequences, covering various weather, illumination, and scenes. All these images are separated from test sequences. To keep the cleanness, we only show recommended coefficients for S_1 in Fig. 4 and briefly visualize all scoring function curves for an intuitive understanding, as Fig. 5. It should be noticed that we provide recommended coefficient values for general purposes. It is often possible to obtain more suitable scoring coefficients with tailored sequences for specific tasks by repeating the aforementioned process. We publish all exact coefficients, source codes, and more information on our website (https://gaozhinuswhu.com).

For final challenge judgment, we calculate the overall perception score S_p by setting $\alpha = 0.4$, $\beta = 0.4$, and $\gamma = 0.2$ to reflect different influences on SLAM according to our common experience. It should be noted that these coefficients can be further adjusted with your customized needs for better performance. Finally, all scores are normalized to [0, 100], and we set the $T_h = 70$ (the center of Good level) and $T_l = 50$ (the center of Fair level) for S_p and other scores, respectively. The CEMS is implemented by Python currently. Each score is defined as an independent function for modular calls. The module only depends on the OpenCV and NumPy library, where we use OpenCV for common computer vision needs and NumPy for efficient matrix operations.

4 Experiments

We conduct extent experiments to demonstrate the feasibility and effectiveness of the proposed method on collected datasets. We first introduce the evaluation datasets (Section 4.1), adopted metrics, and compared methods (Section 4.2). Then, we show quantitative results with a comprehensive analysis (Section 4.3). Finally, we give two initial application demos of the proposed method to exhibit great potentials (Section 4.4).

4.1 Evaluation Datasets

We download EuRoC [54], TUM-RGBD [55], KITTI [56], and AQUALOC [57] dataset to cover the scene of air (UAV), ground (handheld and car), and underwater (UUV), as Fig. 6 shows. For further robustness evaluation, we intentionally set certain frames to pure black or sudden illumination change to mimic the extremely challenging conditions in practical applications, such as the sudden blindness when drones pass through a puff of smoke in rescue tasks and the interruption of image stream due to camera or transmission failure. Specifically, we select the high-quality EuRoC and KITTI datasets as the bases and synthesize visual challenges with automatic Python scripts, generating EuRoC-Syn and KITTI-Syn datasets. Besides public datasets, we collect simulated sequences of a European town with a car in the AirSimNH environment that is officially provided in the AirSim software [58] (supported by Unreal Engine) and many views of Liyue harbor in the Genshin Impact game [59] (supported by Unity Engine). Moreover, we additionally collect sixty sequences in the Genshin Impact game and estimate frame-wise ground truth with ColMap software [60, 61]. The sequences cover a wide range of scenes that are very hard or dangerous to collect, such as the desert, the jungle, and snow mountains, as Fig. 7 shows. Readers may refer to [62] for more detailed information about our Genshin Impact dataset (GID). Moreover, we also obtain a dark tunnel sequence in Wuhan University by field collection with our GZ-LVI device as shown in Fig. 6. The datasets we adopt for test cover a wide range of data sources (from the real world to video games), trajectory lengths (from static view to over 3km displacement), and various challenges (from weather change to adverse illumination).

4.2 Metrics and Compared Methods

We compare differences between calculated and annotated results (overall perception scores S_p and levels) for the con-

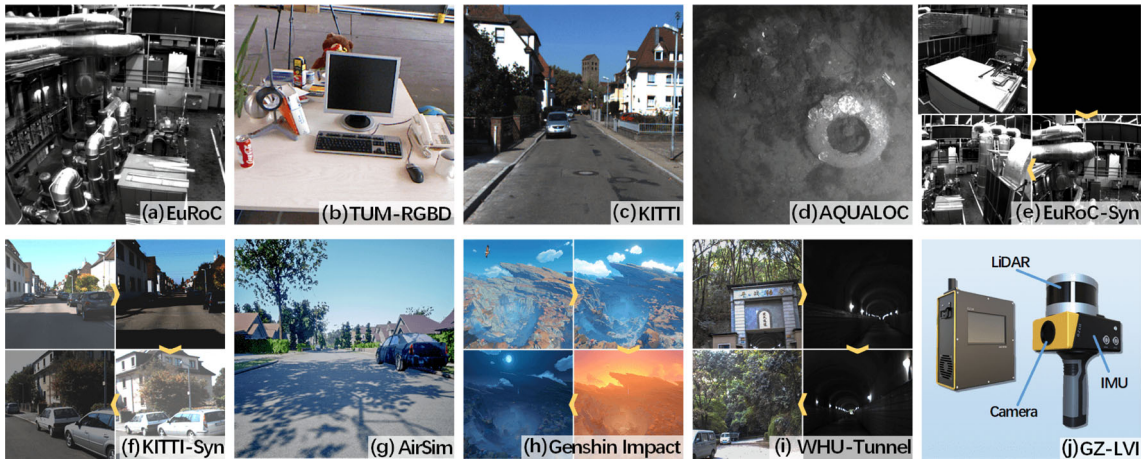

Fig. 6 Sequences come from downloading, synthesis, simulation, and field collection for scoring function estimation and performance tests. (a)-(d) are public datasets, (e) and (f) are synthesized datasets on EuRoC and KITTI. (g) and (h) are simulated sequences in AirSim software and the Genshin Impact game. (i) is the tunnel sequence collected by our (j) GZ-LVI device

sistency evaluation, where the min ($D_{\min}$), max ($D_{\max}$), and mean (D_{mean}) difference are adopted. We also adopt absolute PCC (Pearson Correlation Coefficient) value [63] to measure the correlation between the estimated perception scores and the pose estimation accuracies, which are calculated with commonly used ATE metric [64]. For compared methods, we mainly focus on the tracking performance and select several visual odometry and SLAM methods from traditional methods to learning-based ones for comprehensive experiments, including DSO (direct odometry without global optimization) [65], SVO (semi-direct odometry without global optimization) [66], ORB-SLAM3 (indirect SLAM with global optimization and loop closing) [67] and DROID-SLAM (learning-based end-to-end SLAM with global optimization) [68].

4.3 Results and Analysis

We evaluate the performance of the proposed method from two aspects. First, we compare evaluation results with manual annotation ground truth to demonstrate the feasibility (Section 4.3.1). Second, we calculate the correlation between evaluation results and SLAM pose errors to further indicate the effectiveness (Section 4.3.2).

4.3.1 The Consistency of Evaluation and Challenge Annotation

We obtain the overall perception scores S_p and levels with the proposed module for each frame in all collected datasets and compare them with manual annotations, where the latter is regarded as the ground truth. It is worth noting that all these annotations are based on challenges for SLAM rather than for human eyes, as aforementioned. The range of $D_{\min}$, $D_{\max}$, D_{mean} is [0, 100] for perception scores and [0, 4] for perception levels theoretically, where the smaller suggests the better consistency. These indicators on different datasets are summarized in Table 1. We achieve a mean difference D_{mean} of 11.702 for overall perception score $S_p \in [0, 100]$ and 0.879 for five perception levels, indicating the mean consistency of $100 \times (1 - 11.702/(100 - 0)) = 88.298\%$ and $100 \times (1 - 0.879/(4 - 0)) = 78.025\%$, respectively. These results suggest a good generalization and consistency of our challenge evaluation module on all test datasets, although they come from various sources and scenes. For example, we achieve consistent performance with small (less than about 10% of the range) and similar (10.608 and 7.149) mean difference D_{mean} despite the considerable domain gap between the underwater dataset AQUALOC and the simulation dataset Genshin Impact (Fig. 6). Challenge evaluation results of one frame in the Genshin Impact dataset are shown in Fig. 8. The overall evaluation is challenging for SLAM despite the visually beautiful sunset scenery. Specifically, the S_p is 75.28, higher than the threshold T_h. However, the $S_2 = 34.51$ (dynamic range), $S_5 = 45.89$ (low/rich texture), and $S_7 = 46.57$ (repetitive texture) are lower than threshold T_l, sharing 67.26%, 17.85% 14.89% contribution to the challenge state, respectively. The sunset makes the image orange-toned and significantly reduces the contrast in this scene, bringing difficulties for effective visual tracking. Meanwhile, some similar-appearance trees and no-texture sky also exist, which may exacerbate the challenge. For the whole sequence, we get the frame-wise evaluation results iteratively, and the overall judgment of this sequence is the average of corresponding scores. These evaluation results are of great value for targeted solutions in various applications.

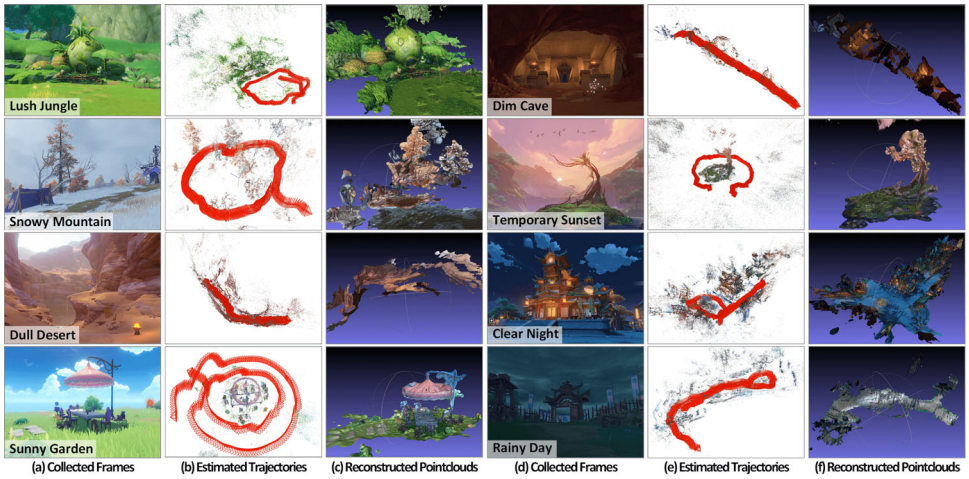

Fig. 7 Some sequences in our Genshin Impact Dataset (GID), covering a wide range of challenging scenes that are rare or dangerous for real-world collection

Moreover, as a widely adopted dataset in the SLAM field, we focus on the EuRoC and compare the evaluation results of all sequences by our module with claimed difficulty levels [54], as summarized in Table 2. The ranking of the mean perception score is consistent with claimed difficulty levels, where scores of hard levels are lower than 84.5, medium levels are among [84.5, 87.5], and easy levels are higher than 87.5. This consistency with EuRoC dataset further demonstrates the feasibility and effectiveness of our method. One more interesting thing we find is that the EuRoC difficulty levels have almost no correlation with max score and hard percent indicators, which stands for the best perception quality and the percentage of challenging frames in the whole sequence, respectively. This finding may reveal the quantitative reasons behind the EuRoC difficulty levels, which are not explicitly explained in their paper [54]. These evaluation results may provide more insights and views to explain tracking performances of existing SLAM methods also evaluated on EuRoC dataset.

4.3.2 The Correlation of Evaluation and Pose Estimation

Due to the high-quality and evenly distributed challenges from easy to hard, we further run selected methods (monocular) on the EuRoC-Syn dataset with pose ground truth and compare frame-wise estimation accuracy with challenge evaluation results, summarized in Table 3. Generally, ORB-SLAM achieves the best ATE of $0.016m$, and the corresponding absolute PCC is 0.944, which suggests the high correlation between the overall perception score S_p and pose estimation errors in SLAM. Moreover, the mean PCC of all test methods is 0.879, demonstrating good generalization and adaptability to various algorithms, including direct and indirect methods, even end-to-end ones.

Moreover, we conduct a dependence analysis on the evaluation results of all collected datasets to demonstrate the effectiveness of proposed scoring functions. First, from the theoretical definition (Eqs. 2 - Eq. 11) and visualization (Fig. 5) of ten scoring functions in CEMS, we can see that all

Table 1 The consistency comparison between evaluation results and manual annotations (Ground truth) on all selected datasets, where the proposed CEMS achieves a high performance

Datasets and sequences	Perception score			Perception level		
	$D_{\min}$	$D_{\max}$	D_{mean}	$D_{\min}$	$D_{\max}$	D_{mean}
EuRoC	0.026	22.235	14.099	0	2	0.801
TUM-RGBD	0.009	25.432	11.802	0	2	1.189
KITTI	0.001	16.268	11.228	0	1	0.911
AQUALOC	0.003	20.162	10.608	0	1	0.524
EuRoC-Syn	0.023	23.613	15.177	0	2	1.031
KITTI-Syn	0.001	26.574	14.978	0	1	1.103
AirSim	0.042	14.407	8.037	0	1	0.646
Genshin Impact	0.066	12.643	7.149	0	1	0.735
WHU-Tunnel	0.021	19.547	12.243	0	1	0.976
Overall	0.021	20.097	11.702	0	1.33	0.879

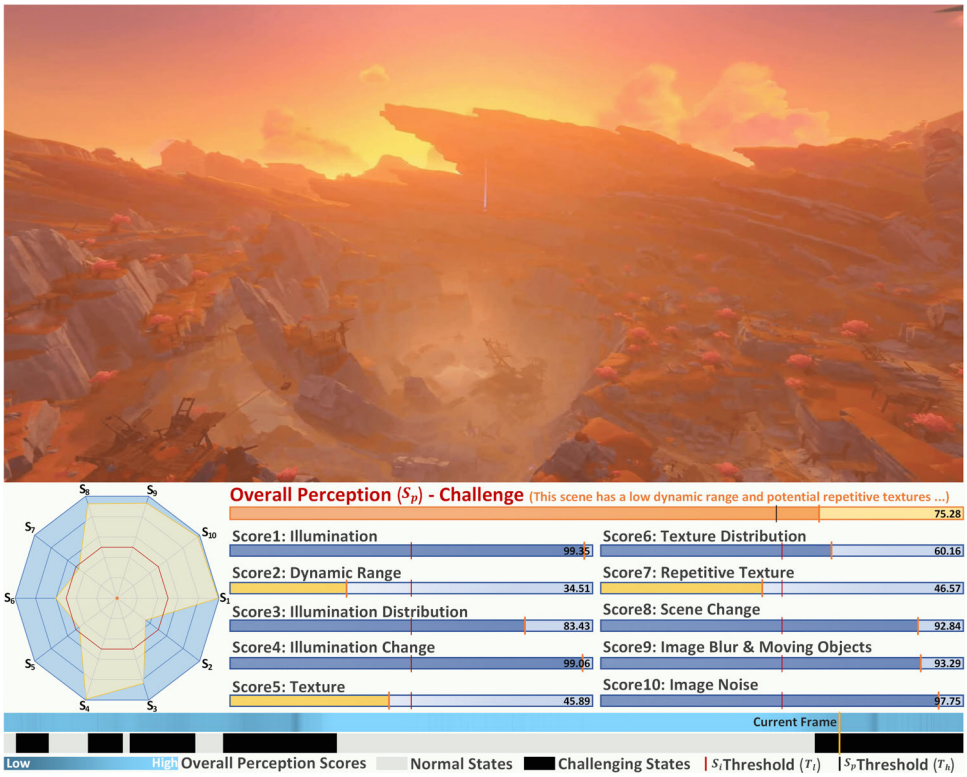

Fig. 8 Evaluation results of one frame in our Genshin Impact sequences. Ten scores S_i are visualized in blue bars and red lines indicate corresponding threshold T_l, where scores below T_l are marked in yellow and others are in dark blue. The overall score S_p is visualized in an orange bar and T_h is represented with a black line. Judgments are visualized in different colors, where darker colors indicate challenging states in the sequence

Table 2 The comparison of challenge evaluation results evaluated by us and claimed difficulty levels (last column) of EuRoC dataset

Seq.	Hard percent	Max score	Min score	Score variance	Mean score	Claimed level [54]
MH01	23.22%	94.911	67.916	15.554	87.808	Easy
MH02	27.64%	94.801	68.228	17.457	87.816	Easy
MH03	11.30%	94.242	69.445	18.903	87.067	Medium
MH04	22.63%	94.279	38.020	178.649	82.229	Hard
MH05	22.71%	93.581	42.850	160.611	82.416	Hard
VR101	2.23%	92.895	70.012	11.097	87.784	Easy
VR102	11.41%	92.009	66.575	10.857	86.277	Medium
VR103	47.76%	92.282	50.065	54.193	81.857	Hard
VR201	5.44%	93.142	69.560	7.241	88.809	Easy
VR202	9.92%	92.786	68.723	12.489	87.322	Medium
VR203	4.26%	92.283	61.853	22.031	84.182	Hard

Table 3 The correlation between challenge evaluation and pose estimation of different odometry/SLAM on our EuRoC-Syn sequence

Odometry/SLAM	Perception score of EuRoC-Syn Seq.			Frame-wise Estimation Error (Unit: m)			Abs. PCC
	Max	Min	Mean	Max	Min	Mean	
ORB-SLAM3 (SLAM) [67]				0.153	0.001	0.016	0.944
DSO (Odometry) [65]				0.067	0.006	0.027	0.925
SVO (Odometry) [66]	90.158	16.512	81.116	0.144	0.003	0.033	0.853
DROID-SLAM (SLAM) [68]				0.117	0.007	0.066	0.795
Mean				0.012	0.004	0.036	0.879

these scores are generally independent of others since each function has its unique input with different trends, representing various evaluation metric. For example, S_1 has a quadratic function-like curve with the mean grayscale input while S_5 has a negative power function-like trend with grayscale gradient as input. Therefore, they are not correlated to each other. Despite similar trends in some functions (such as S_2, S_6, S_8, and S_9), they have totally different inputs, resulting in weak correlations between each other. Second, to further demonstrate the independence of scoring functions, we conduct quantitative correlation analysis on each score with PCC. We calculate the absolute PCC among each score pair in test sequences and obtain many PCC matrices. Then, we average these PCC matrices to get the overall final PCC matrix, as Fig. 9(a) shows. It can be seen that for any score, it generally has a weak correlation with other scores. Moreover, we divide the PCC into four categories according to the value, including none correlation (PCC $\leq$ 0.3), weak correlation (0.3 $<$ PCC $\leq$ 0.5), moderate correlation (0.5 $<$ PCC $\leq$ 0.8), and high correlation (PCC $>$ 0.8) [69]. Generally, as Fig. 9(b) shows, any score in CEMS typically has six or more scores that have no or weak correlation with it. Both theoretical and quantitative analysis suggest the effectiveness of the proposed scoring functions and CEMS module.

4.3.3 Validation of Computational Efficiency

Generally, the proposed challenge evaluation module is mainly for practical and real-time robot tasks (such as search and rescue in adverse scenes) rather than the merely offline dataset evaluation. We want to achieve an efficient and quantitative evaluation of visual challenges before tracking, bringing more potential for tailored enhancement solutions. Therefore, besides the feasibility and effectiveness of the proposed method, the computational efficiency is also of our consideration since computation resource is usually very limited in many mobile platforms. The CEMS should not occupy much computation load and leave valuable power and computation resources for other tasks such as motion planning and controlling.

We reorganize all selected datasets according to frame resolution and conduct experiments for computational efficiency on various platforms. We evaluate the efficiency of our module on various platforms, including Raspberry Pi 4B+, Nvidia Jetson AGX Xavier, Dell XPS-15 Laptop (Intel i7-10875H CPU), and ASUS Workstation (Intel i9-9900K CPU). We select several typical frame resolutions to simulate possible conditions in the real world, such as the 640 $\times$ 480 size. We statistically calculate the cost time (in seconds) for evaluation by inserting timing codes and obtain the final cost time with the averaging of five funs, as summarized in Table 4. It can be seen that the module runs rapidly on different platforms. For example, despite the initial implementation with Python, the module runs stably at 20 FPS (frames per second) on our workstation with 640 $\times$ 480 resolution. Even on the Raspberry Pi with very limited computational resources, we can also achieve around 4 FPS with 640 $\times$ 480 resolution. Therefore, the proposed method can generally run in real-time for practical tasks with adverse conditions. Moreover, since cameras generally have a high FPS of around 20 to 30 with small frame-wise difference. It is often not necessary to evaluate every input frame in practical applications. Similar to the widely adopted key frame strategy in SLAM, we can also conduct the quantitative evaluation every several frames to further decrease computation for cameras with very high frequency. For example, according to experimental results, we achieved around 4 FPS on the Raspberry platform. Thus, we can parallelly evaluate challenges every seven or eight frames, handling a camera input stream with around 28 or 32 FPS (640x480). Finally, it also should be noticed that the efficiency can be improved by implementation with C++ language.

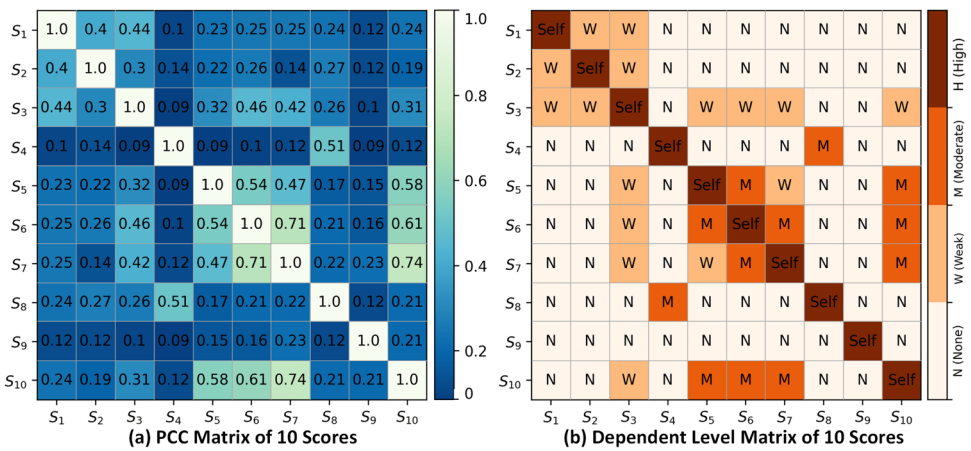

Fig. 9 PCC and dependent matrix of 10 scores in CEMS. (a) PCC matrix between scores, where the darker color indicates less correlation. (b) Dependent levels between scores, where the dependence is divided into four levels

Table 4 Time cost for single frame evaluation of our module on various platforms with different frame resolutions (unit: second)

Frame resolution	Raspberry Pi 4B+	Jetson AGX Xavier	Dell laptop	ASUS workstation
640×480 (30MP)	0.294	0.095	0.065	0.050
968×608 (60MP)	0.572	0.162	0.078	0.059
1600×900 (140MP)	1.359	0.338	0.149	0.109
1920×1080 (200MP)	1.768	0.460	0.223	0.166

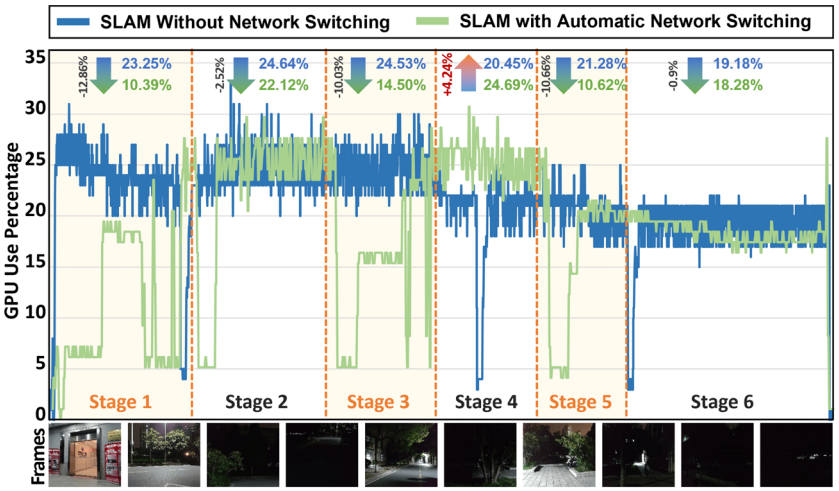

Fig. 10 Comparison of computational efficiency (evaluated by GPU load) between SLAM without switching strategy (blue line) and SLAM with switching strategy based on CEMS (green line). In stage 1,3,5, the neural network is turned off due to good illumination, resulting in a significant decrease of GPU load. While in stage 2,4,6, the network is automatically turned on for enhancement

Table 5 Overall evaluation of various challenges in representative SLAM datasets (in descending order of challenges)

SLAM dataset	Overall perception Score	Biggest challenge	Corresponding percentage
ICL-NUIM [48]	75.567	Texture distribution	59.54%
AQUALOC [57]	76.778	Texture distribution	66.05%
TartanAir [70]	80.369	Low illumination	26.75%
FusionPortable [71]	81.183	Texture distribution	24.67%
Lyft5 [72]	82.257	Repetitive texture	14.29%
nuScenes [73]	82.344	Low illumination	8.87%
KITTI [56]	85.000	Texture distribution	0.67%
EuRoC [54]	86.075	Low illumination	11.04%

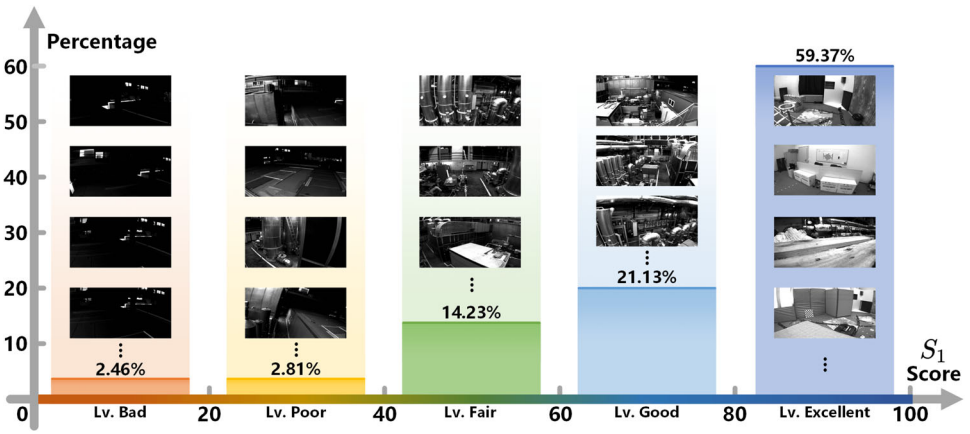

Fig. 11 Visualization for illumination scores (S_1) and the distribution of the percentage at different perception levels in EuRoC dataset, where level excellent accounts most in the dataset

4.4 Potential Applications of CEMS

In this section, we show some primary and valuable applications of CEMS for practical SLAM-related tasks, which mainly involve the online aid for robust tracking and the offline evaluation for dataset benchmarking.

4.4.1 Robust SLAM with Degradation Perception Towards Adverse Illumination

As aforementioned, bad illumination is one of the biggest challenges in robust exploration for SLAM. However, current SLAM either cannot handle adverse conditions or struggles with constant heavy computations (such as neural networks). It is usually not affordable on practical mobile platforms. Following the aforementioned idea of "enhance on-demand", we initially developed a visual SLAM that can perceive ambient illumination changes and react with optimal strategies, where the CEMS module plays an important role in condition switching. If the evaluation result is poor, we will leverage a neural network for illumination enhancement; otherwise, we will mute the network for higher efficiency and power saving. Benefiting from the proposed CEMS, the prototype visual SLAM is robust to adverse illumination while keeping high computational efficiency.

We primarily test the SLAM with a field-collected sequence taken on the campus of Wuhan University at night. The computation burden of the device is shown in Fig. 10. With the proposed switching strategy, we significantly reduce the workload of GPU when illumination is good (Stage 1,3,5 marked in orange) and automatically turn on the network when the system encounters challenging environments (Stage 2,4,6 marked in white). It is worth noting that this improvement could be more significant if the whole sequence is in good condition, which suggests that we turn off networks rather than enhance good frames all the time. We will continue working on it and make improvements with more comprehensive analysis and experiments.

4.4.2 Quantitative Challenge Evaluation for Representative SLAM Datasets

Besides the real-time applications in SLAM, we can also use the CEMS independently to evaluate various datasets for a fast and brief grasp before using them. As an initial introduction to the following work, we comprehensively evaluate challenges over 1,450,000 frames in public SLAM datasets with our module and generate the CET (Challenge Evaluation Table). Every dataset is evaluated automatically with detailed scores and overall reports. The overall perception score is the mean of every perception score of sequences in the dataset. Some overall results are summarized in Table 5.

Generally, ICL-NUIM is the most challenging among evaluated datasets with the lowest overall perception score. The biggest challenge in ICL-NUIM is the uneven-distributed textures (59.54%). While EuRoC achieves the highest overall perception score of 86.075, suggesting its high quality and reasons for wide adoption in the SLAM field. Due to limited space in this paper, we briefly show the distribution of illumination score S_1 and corresponding levels in EuRoC dataset in Fig. 11. It can be seen that excellent illumination accounts for most in the dataset (around 59.37%), while only nearly 5% are in bad illumination. More information and analysis on SLAM datasets can be found on our website https://gaozhinuswhu.com. We will continue working to develop and maintain webpages with integrated and convenient functions that are accessible to everyone to benefit the society, including but not limited to online evaluation, score comparison, result analysis, report generation, and challenge benchmarking.

5 Conclusions

We focus on challenging environments for visual SLAM and propose an innovative evaluation module CEMS for the automatic degradation awareness of unmanned systems. Extensive experiments on various datasets demonstrate the effectiveness of our method. Moreover, we build CET, the first table for the quantitative evaluation of challenges in various SLAM datasets. To our best knowledge, there are no similar works at present. For future works, we will continually refine the analysis framework (including an adaptive scheme for coefficients based on fuzzy logic) and improve the evaluation module with C++ and CUDA for higher efficiency. Moreover, we will extend our analysis framework and CEMS to other data sources to cover more SLAM frameworks, including the thermal and multi/hyperspectral images, even stereo disparity images. We will also enlarge the CET with more data and develop scripts and websites to benefit the community.

Acknowledgements This research is partially supported by Wuhan University - Huawei Geoinformatics Innovation Laboratory. Some numerical calculations in this paper have been supported by the supercomputing system in the Supercomputing Center of Wuhan University.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Xuhui Zhao and Zhi Gao. The first draft of the manuscript was written by Xuhui Zhao and all authors commented on previous versions of the manuscript. All authors read and approved the fnal manuscript.

Funding This research is partially supported by the National Natural Science Foundation of China Major Program (Grant No. 42192580, 42192583), Hubei Province Natural Science Foundation (Grant No. 2021CFA088 and 2020CFA003), and the Science and Technology Major Project (Grant No. 2021AAA010, 2021AAA010-3).

Data Availability All data and materials generated or analysed during this study are included in this published article and its supplementary information files.

Code Availability The code generated during the current study will be refined and then be available on GitHub.

Declarations

Conflict of Interest The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval The authors declare that no human or animal subjects are involved in the study.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication Patients signed informed consent regarding publishing their data and photographs.

References

1. Chen, B.M.: On the trends of autonomous unmanned systems research. Engineering **12**, 20–23 (2021)
2. Bujanca, M., Shi, X., Spear, M., Zhao, P., Lennox, B., Luján, M.: Robust slam systems: are we there yet? In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5320–5327 (2021)
3. Garforth, J., Webb, B.: Visual appearance analysis of forest scenes for monocular slam. In: 2019 International Conference on Robotics and Automation (ICRA), pp. 1794–1800 (2019)
4. Park, S., Schöps, T., Pollefeys, M.: Illumination change robustness in direct visual slam. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 4523–4530 (2017)
5. CVPR 2020 SLAM Challenge. https://sites.google.com/view/vislocslamcvpr2020/slam-challenge
6. Liu, X., Gao, Z., Chen, B.M.: Ipmgan: integrating physical model and generative adversarial network for underwater image enhancement. Neurocomputing **453**, 538–551 (2021)
7. Rahman, S., Li, A.Q., Rekleitis, I.: Svin2: an underwater slam system using sonar, visual, inertial, and depth sensor. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1861–1868 (2019)
8. Zhou, L., Huang, G., Mao, Y., Wang, S., Kaess, M.: Edplvo: efficient direct point-line visual odometry. In: 2022 International Conference on Robotics and Automation, pp. 7559–7565 (2022)
9. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: self-supervised interest point detection and description. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 337–33712 (2018)
10. Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: learning feature matching with graph neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4937–4946 (2020)
11. Joo, K., Oh, T.-H., Kweon, I.S., Bazin, J.-C.: Globally optimal inlier set maximization for atlanta world understanding. IEEE Trans. Pattern Anal. Mach. Intell. **42**(10), 2656–2669 (2020)
12. Yunus, R., Li, Y., Tombari, F.: Manhattanslam: Robust planar tracking and mapping leveraging mixture of manhattan frames. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 6687–6693 (2021)
13. Qiu, Y., Wang, C., Wang, W., Henein, M., Scherer, S.: Airdos: dynamic slam benefits from articulated objects. In: 2022 International Conference on Robotics and Automation, pp. 8047–8053 (2022)
14. Tomasi, J., Wagstaff, B., Waslander, S.L., Kelly, J.: Learned camera gain and exposure control for improved visual feature detection and matching. IEEE Robotics and Automation Letters **6**(2), 2028–2035 (2021)
15. Brunner, C., Peynot, T., Underwood, J.: Towards discrimination of challenging conditions for ugvs with visual and infrared sensors. In: ARAA Australasian Conference on Robotics and Automation, Sydney, Australia (2009)
16. Brunner, C., Peynot, T.: Visual metrics for the evaluation of sensor data quality in outdoor perception. In: Proceedings of the 10th Performance Metrics for Intelligent Systems Workshop, pp. 1–8 (2010)
17. Brunner, C., Peynot, T., Vidal-Calleja, T.: Combining multiple sensor modalities for a localisation robust to smoke. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2489–2496 (2011)
18. Brunner, C., Peynot, T., Vidal-Calleja, T., Underwood, J.: Selective combination of visual and thermal imaging for resilient localization in adverse conditions: day and night, smoke and fire. Journal of Field Robotics **30**(4), 641–666 (2013)
19. Brunner, C., Peynot, T.: Perception quality evaluation with visual and infrared cameras in challenging environmental conditions. In: Experimental Robotics: The 12th International Symposium on Experimental Robotics, pp. 711–725 (2014). Springer
20. Kim, P., Coltin, B., Alexandrov, O., Kim, H.J.: Robust visual localization in changing lighting conditions. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 5447–5452 (2017)
21. DARPA Subterranean(SubT) Challenge. www.darpa.mil/program/darpa-subterranean-challenge
22. Tranzatto, M., Miki, T., Dharmadhikari, M., Bernreiter, L., Kulkarni, M., Mascarich, F., Andersson, O., Khattak, S., Hutter, M., Siegwart, R., et al.: Cerberus in the darpa subterranean challenge. Sci. Robot. **7**(66), 9742 (2022)
23. Carrillo, H., Reid, I., Castellanos, J.A.: On the comparison of uncertainty criteria for active slam. In: 2012 IEEE International Conference on Robotics and Automation, pp. 2080–2087 (2012)
24. Agha, A., Otsu, K., Morrell, B., Fan, D.D., Thakker, R., Santamaria-Navarro, A., Kim, S.-K., Bouman, A., Lei, X., Edlund, J., et al.: Nebula: quest for robotic autonomy in challenging environments; team costar at the darpa subterranean challenge. (2021). arXiv:2103.11470
25. Santamaria-Navarro, A., Thakker, R., Fan, D.D., Morrell, B., Aghamohammadi, A.-a.: Towards resilient autonomous navigation of drones. In: Robotics Research: The 19th International Symposium ISRR, pp. 922–937 (2022). Springer

26. Kramer, A., Stahoviak, C., Santamaria-Navarro, A., Aghamohammadi, A.-A., Heckman, C.: Radar-inertial ego-velocity estimation for visually degraded environments. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 5739–5746 (2020). IEEE
27. Palieri, M., Morrell, B., Thakur, A., Ebadi, K., Nash, J., Chatterjee, A., Kanellakis, C., Carlone, L., Guaragnella, C., Agha-mohammadi, A.-a.: Locus: a multi-sensor lidar-centric solution for high-precision odometry and 3d mapping in real-time. IEEE Robotics and Automation Letters 6(2), 421–428 (2021)
28. Tagliabue, A., Tordesillas, J., Cai, X., Santamaria-Navarro, A., How, J.P., Carlone, L., Agha-mohammadi, A.-a.: Lion: Lidar-inertial observability-aware navigator for vision-denied environments. In: Experimental Robotics: The 17th International Symposium, pp. 380–390 (2021). Springer
29. Ebadi, K., Chang, Y., Palieri, M., Stephens, A., Hatteland, A., Heiden, E., Thakur, A., Funabiki, N., Morrell, B., Wood, S., Carlone, L., Agha-mohammadi, A.-a.: Lamp: large-scale autonomous mapping and positioning for exploration of perceptually-degraded subterranean environments. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 80–86 (2020)
30. Ebadi, K., Palieri, M., Wood, S., Padgett, C., Agha-mohammadi, A.-a.: Dare-slam: degeneracy-aware and resilient loop closing in perceptually-degraded environments. Journal of Intelligent & Robotic Systems 102, 1–25 (2021)
31. Rouček, T., Pecka, M., Cížek, P., Petříček, T., Bayer, J., Šalanskì, V., Heřt, D., Petrlík, M., Báča, T., Spurnỳ, V., et al.: Darpa subterranean challenge: multi-robotic exploration of underground environments. In: Modelling and Simulation for Autonomous Systems: 6th International Conference, MESAS 2019, Palermo, Italy, October 29–31, 2019, Revised Selected Papers 6, pp. 274–290 (2020). Springer
32. Zhang, L., Zhang, L., Mou, X., Zhang, D.: Fsim: a feature similarity index for image quality assessment. IEEE Trans. Image Process. **20**(8), 2378–2386 (2011)
33. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans. Image Process. **20**(12), 3350–3364 (2011)
34. Ma, K., Liu, W., Zhang, K., Duanmu, Z., Wang, Z., Zuo, W.: End-to-end blind image quality assessment using deep neural networks. IEEE Trans. Image Process. **27**(3), 1202–1213 (2018)
35. Zhu, H., Li, L., Wu, J., Dong, W., Shi, G.: Metaiqa: deep meta-learning for no reference image quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14143–14152 (2020)
36. Cheon, M., Yoon, S.-J., Kang, B., Lee, J.: Perceptual image quality assessment with transformers. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 433–442 (2021)
37. Yang, N., Zhong, Q., Li, K., Cong, R., Zhao, Y., Kwong, S.: A reference-free underwater image quality assessment metric in frequency domain. Signal Processing: Image Communication **94**, 116218 (2021)
38. Xiang, T., Yang, Y., Guo, S.: Blind night-time image quality assessment: subjective and objective approaches. IEEE Trans. Multimedia **22**(5), 1259–1272 (2020)
39. Liu, W., Zhou, F., Lu, T., Duan, J., Qiu, G.: Image defogging quality assessment: real-world database and method. IEEE Trans. Image Process. **30**, 176–190 (2021)
40. Li, X.: Blind image quality assessment. In: 2002 IEEE International Conference on Image Processing, vol. 1, p. (2002)
41. Mier, J.C., Huang, E., Talebi, H., Yang, F., Milanfar, P.: Deep perceptual image quality assessment for compression. In: 2021 IEEE International Conference on Image Processing, pp. 1484–1488 (2021)
42. Ma, K., Zeng, K., Wang, Z.: Perceptual quality assessment for multi-exposure image fusion. IEEE Trans. Image Process. **24**(11), 3345–3356 (2015)
43. Dendi, S.V.R., Channappayya, S.S.: No-reference video quality assessment using natural spatiotemporal scene statistics. IEEE Trans. Image Process. **29**, 5612–5624 (2020)
44. Zhang, J., Kaess, M., Singh, S.: On degeneracy of optimization-based state estimation problems. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 809–816 (2016)
45. Zhang, J., Singh, S.: Enabling aggressive motion estimation at low-drift and accurate mapping in real-time. In: IEEE International Conference on Robotics and Automation, pp. 5051–5058 (2017)
46. Thakker, R., Alatur, N., Fan, D.D., Tordesillas, J., Paton, M., Otsu, K., Toupet, O., Agha-mohammadi, A.-a.: Autonomous off-road navigation over extreme terrains with perceptually-challenging conditions. In: Experimental Robotics: The 17th International Symposium, pp. 161–173 (2021). Springer
47. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer, Cham (2022)
48. Handa, A., Whelan, T., McDonald, J., Davison, A.J.: A benchmark for rgb-d visual odometry, 3d reconstruction and slam. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 1524–1531 (2014)
49. Cepeda-Negrete, J., Sanchez-Yanez, R.E.: Gray-world assumption on perceptual color spaces. In: Image and Video Technology: 6th Pacific-Rim Symposium, PSIVT 2013, Guanajuato, Mexico, October 28-November 1, 2013. Proceedings 6, pp. 493–504 (2014). Springer
50. Tranzatto, M., Mascarich, F., Bernreiter, L., Godinho, C., Camurri, M., Khattak, S., Dang, T., Reijgwart, V., Loeje, J., Wisth, D.: Cerberus: autonomous legged and aerial robotic exploration in the tunnel and urban circuits of the darpa subterranean challenge. (2022). arXiv:2201.07067
51. Mur-Artal, R., Montiel, J.M.M., Tardós, J.D.: Orb-slam: a versatile and accurate monocular slam system. IEEE Trans. Rob. **31**(5), 1147–1163 (2015)
52. Gadkari, D.: Image quality analysis using glcm (2004)
53. BT, I.: Methodologies for the subjective assessment of the quality of television images, document recommendation itu-r bt. 500–14 (10/2019). ITU, Geneva, Switzerland (2020)
54. Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., Achtelik, M.W., Siegwart, R.: The euroc micro aerial vehicle datasets. The International Journal of Robotics Research **35**(10), 1157–1163 (2016)
55. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of rgb-d slam systems. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 573–580 (2012)
56. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: the kitti dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)
57. Ferrera, M., Creuze, V., Moras, J., Trouvé-Peloux, P.: Aqualoc: an underwater dataset for visual–inertial–pressure localization. The International Journal of Robotics Research **38**(14), 1549–1559 (2019)
58. Shah, S., Dey, D., Lovett, C., Kapoor, A.: Airsim: high-fidelity visual and physical simulation for autonomous vehicles. In: International Symposium on Field and Service Robotics (2017)
59. HoYoverse: Genshin Impact-Step Into a Vast Magical World of Advanture. (2023). https://genshin.hoyoverse.com/en
60. Schönberger, J.L., Frahm, J.-M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
61. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.-M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV) (2016)
62. Zhao, X.: The Genshin Impact Dataset (GID) for SLAM. https://github.com/zhaoxuhui/Genshin-Impact-Dataset

63. Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., Huang, Y., Cohen, I.: Pearson correlation coefficient. Noise reduction in speech processing, 1–4 (2009)
64. Zhang, Z., Scaramuzza, D.: A tutorial on quantitative trajectory evaluation for visual(-inertial) odometry. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7244–7251 (2018). IEEE
65. Engel, J., Koltun, V., Cremers, D.: Direct sparse odometry. IEEE Trans. Pattern Anal. Mach. Intell. **40**(3), 611–625 (2018)
66. Forster, C., Pizzoli, M., Scaramuzza, D.: Svo: fast semi-direct monocular visual odometry. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 15–22 (2014)
67. Campos, C., Elvira, R., Rodríguez, J.J.G., M. Montiel, J.M., D. Tardós, J.: Orb-slam3: an accurate open-source library for visual, visual–inertial, and multimap slam. IEEE Transactions on Robotics 37(6), 1874–1890 (2021)
68. Teed, Z., Deng, J.: Droid-slam: deep visual slam for monocular, stereo, and rgb-d cameras. Adv. Neural. Inf. Process. Syst. **34**, 16558–16569 (2021)
69. Moore, D.S.: Statistics: Concepts and controversies. (1980)
70. Wang, W., Zhu, D., Wang, X., Hu, Y., Qiu, Y., Wang, C., Hu, Y., Kapoor, A., Scherer, S.: Tartanair: A dataset to push the limits of visual slam. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4909–4916 (2020)
71. Jiao, J., Wei, H., Hu, T., Hu, X., Zhu, Y., He, Z., Wu, J., Yu, J., Xie, X., Huang, H., Geng, R., Wang, L., Liu, M.: Fusionportable: a multi-sensor campus-scene dataset for evaluation of localization and mapping accuracy on diverse platforms, 3851–3856 (2022)
72. Houston, J., Zuidhof, G., Bergamini, L., Ye, Y., Chen, L., Jain, A., Omari, S., Iglovikov, V., Ondruska, P.: One thousand and one hours: self-driving motion prediction dataset. In: Conference on Robot Learning, pp. 409–418 (2021). PMLR
73. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: Nuscenes: a multimodal dataset for autonomous driving. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11618–11628 (2020)

Xuhui Zhao received the master's degree from State the Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, China. He is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering. He has published more than ten academic articles in prestigious conferences and journals, such as IEEE IROS, IEEE IGARSS, ISPRS Congress, IEEE Transactions on Intelligent Transportation Systems, and Journal of Intelligent & Robotic Systems. He is the Program Session Chair of IEEE IGARSS. His research interests include visual perception for unmanned systems, SLAM, low rank analysis, satellite video understanding, and robotics.

Zhi Gao received the B.E. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2002 and 2007, respectively. In 2008, he joined the Interactive and Digital Media Institute, National University of Singapore (NUS), as a Research Fellow and the Project Manager. In 2014, he joined the Temasek Laboratories, NUS (TL@NUS), as a Research Scientist and the Principal Investigator. He is currently a Full Professor with the School of Remote Sensing and Information Engineering, Wuhan University. He has published more than 90 academic articles, which have been published in International Journal of Computer Vision, IEEE Transactions on Pattern Analysis and Machine Intelligence, ISPRS Journal of Photogrammetry and Remote Sensing, IEEE Transactions on Geoscience and Remote Sensing, IEEE Transactions on Intelligent Transportation System, and other top journals. He received the prestigious "National Plan for Young Talents" Award and the Hubei Province Funds for Distinguished Young Scientists. In addition, he is a "Chutian Scholar" Distinguished Professor in Hubei. He serves as an Associate Editor for the journal Unmanned Systems.

Hao Li received his bachelor degree in Electronics Engineering from the Electronic Information School, Wuhan University. He is currently a Ph. D. candidate in the School of Remote Sensing and Information Engineering, Wuhan University. His research is primarily focused on multi-robot collaborative localization and mapping in large-scale complex environments.

Hong Ji received the B.E. and M.S. degrees from the School of Electronic Information, Wuhan University, Wuhan, China, in 2017 and 2020, respectively, where she is currently pursuing the Ph.D. degree with the School of Remote Sensing and Information Engineering. Her research interests include computer vision, artificial intelligence, few-shot learning, and their applications.

Hong Yang received the master's degree in industry and business administration from Beihang University, Beijing, China, in 2010, and is currently working toward the Engineering Doctor degree with the School of Electronic Information, Northwestern Polytechnical University, Xi'an, China. He is currently working with Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include the overall design and application of high-resolution aerial remote sensing.

Chenyang Li received his bachelor's degree from the School of Remote Sensing and Information Engineering, Wuhan University, China. He is currently pursuing the master's degree in the School of Remote Sensing and Information Engineering, Wuhan University. His research interests include LiDAR SLAM and point cloud registration.

Hao Fang received the B.S. degree from the Xi'an University of Technology, Shaanxi, China, in 1995, and the M.S. and Ph.D. degrees from the Xi'an Jiaotong University, Shaanxi, in 1998 and 2002, respectively. Since 2011, he has been a Professor with the Beijing Institute of Technology, Beijing, China. He held two postdoctoral appointments with the INRIA/France Research Group of COPRIN and the LASMEA (UNR6602 CNRS/Blaise Pascal University, Clermont-Ferrand, France). His research interests include all-terrain mobile robots, robotic control, and multiagent systems.

Ben M. Chen (Fellow, IEEE) received the B.Sc. degree in mathematics and computer science from Xiamen University, China, in 1983, the M.Sc. degree in electrical engineering from Gonzaga University, USA, in 1988, and the Ph.D. degree in electrical and computer engineering from Washington State University, USA, in 1991. He was the Provost's Chair Professor with the Department of Electrical and Computer Engineering, National University of Singapore, before joining The Chinese University of Hong Kong (CUHK) in 2018. He was an Assistant Professor with the Department of Electrical Engineering, the State University of New York at Stony Brook, USA, from 1992 to 1993. He is currently a Professor of mechanical and automation engineering with CUHK. He has authored/coauthored more than 500 journal and conference papers, and a dozen research monographs in control theory and applications, unmanned systems, and financial market modeling. His current research interests include unmanned systems and their applications. He is a Fellow of Academy of Engineering, Singapore. He served on the editorial boards for a dozen international journals, including Automatica and IEEE Transactions on Automatic Control. He is serving as the Editor-in-Chief for Unmanned Systems, the Editor for International Journal of Robust and Nonlinear Control, and an Editorial Member of Science China Information Sciences.

Authors and Affiliations

Xuhui Zhao[1] · Zhi Gao[1] · Hao Li[1] · Hong Ji[1] · Hong Yang[2] · Chenyang Li[1] · Hao Fang[3] · Ben M. Chen[4]

Xuhui Zhao
zhaoxuhui@whu.edu.cn

Hao Li
leoli9901@whu.edu.cn

Hong Ji
jihong@whu.edu.cn

Hong Yang
yanghong@aircas.ac.cn

Chenyang Li
2018302130131@whu.edu.cn

Hao Fang
fangh@bit.edu.cn

Ben M. Chen
bmchen@cuhk.edu.hk

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, Hubei, China

2 Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

3 School of Automation, Beijing Institute of Technology, Beijing 100081, China

4 Department of Mechanical and Automation Engineering, Chinese University of Hong Kong, Hong Kong 999077, China