2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

An Image Acquisition Scheme for Visual Odometry based on Image Bracketing and Online Attribute Control

Shuyang Zhang[1], Jinhao He[2], Bohuan Xue[1], Jin Wu[1], Pengyu Yin[3], Jianhao Jiao[1,4*] and Ming Liu[2]

***Abstract*— Visual odometry (VO) system is challenged by complex illumination environments. Image quality and its consistency in the time domain directly determine feature detection and tracking performance, which further affect the robustness and accuracy of the entire system. In this paper, an image acquisition scheme with image bracketing patterns is proposed. Images with different exposure levels are continuously captured to sufficiently explore the scene under varying illumination. An attribute control method is designed to adjust image exposures within the brackets online. Gaussian process regression fits the relationship between image quality metric and exposure via image synthesis technique. The optimal exposures for the next bracket are obtained directly without attempts to ensure a quick response. Experiments show our acquisition system's effectiveness and performance improvement for VO tasks in complex illumination scenes.**

I. Introduction

Camera sensors nowadays are still difficult to perfectly adapt to visual odometry (VO) tasks, especially under complex illumination environments [1]–[3]. One limitation is from camera's dynamic range. Digital cameras with global shutters have a dynamic range of around 70 dB, while high dynamic range (HDR) scenes usually exceed 100 dB. Features are drowned in underexposed and overexposed areas, finally causing an accuracy drop or system failure. Another issue is that the camera acquisition frequency (over 40 Hz) does not match the working frequency of VO systems (commonly at 10–20 Hz [4]–[6]). Higher acquisition frequency only strengthens the stability of inter-frame feature tracking and does not boost the feature detection performance in keyframes, directly influencing state estimation performance.

In addition to introducing new camera sensors [7], [8] or image enhancement techniques [9], [10], two principal image capturing schemes are proposed to improve image quality for VO systems in high dynamic range scenes. One introduces the HDR photography technique, capturing multiple low dynamic range (LDR) images with different exposures and generating an HDR image by tone mapping. In addition to a high computation cost, this approach suffers from the visual artifacts caused by HDR fusion since the cameras of VO systems always move rapidly. Kim *et al.* [11] proposed D-HDR, which generates HDR images with only one low-exposed seed image. Higher exposed images are synthesized from the seed, and artifacts from camera motion are avoided. However, the seed image's quality greatly influences the synthesis performance, and the exposure of the seed needs to be carefully selected according to the scene. Another solution is to adjust the camera's exposure parameters (or attributes) online as the scene lighting changes. Given previous images, the optimal attributes for the next image are estimated. These approaches provide a higher output rate to VO systems than traditional HDR solutions. However, due to the dynamic range gap between the camera and the scene, the scene information is partially observed, and the estimation may be locally optimal. In addition, attribute control methods with feedback control framework have convergence delay in extreme lighting cases and may lead to the performance degradation of VO systems.

Combining these two schemes, we propose a camera acquisition system with online camera attribute control for VO systems under HDR scenes. Our contributions contain:

- A camera attribute control method adapted to image bracketing patterns. Images with various exposures are captured for scene exploration, and optimal exposure for the next control is globally optimized by Gaussian process regression (GPR).
- A VO-oriented image acquisition scheme that explores a wide dynamic range and provides stable image sequences in time domain. The system leverages the exploration of the dynamic range with system output frequency according to the bracketing pattern design.
- Experimental comparisons with various baseline methods, which show the effectiveness and robustness of our method under complex illuminations.

II. Related Work

Camera attribute control methods based on gradient descent are widely used due to their high calculation efficiency. Zhang *et al.* [12] redesigned the image gradient metric for practical use and gave detailed proof of the image gradient derivative w.r.t. their metrics. They also considered the photometric compensation for both direct and feature-based VO systems. Wang *et al.* [13] calculated the derivative of their metric function and proposed a heuristic exposure control algorithm. They added an online photometric calibration approach and a photometric compensation module. Han *et al.* [14] used a gradient-based iterative search framework

[1]S. Zhang, B. Xue, J. Wu, J. Jiao are with the Department of Electronic and Computer Engineering, the Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China. (szhangcy@connect.ust.hk)

[2]J. He and M. Liu are with the Thrust of Robotics and Autonomous Systems, the Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China.

[3]P. Yin is with the Center for Advanced Robotics Technology Innovation (CARTIN), School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

[4]J. Jiao is with the Department of Computer Science, University College London, Gower Street, WC1E 6BT, London, UK. (ucacjji@ucl.ac.uk)

*Corresponding author.

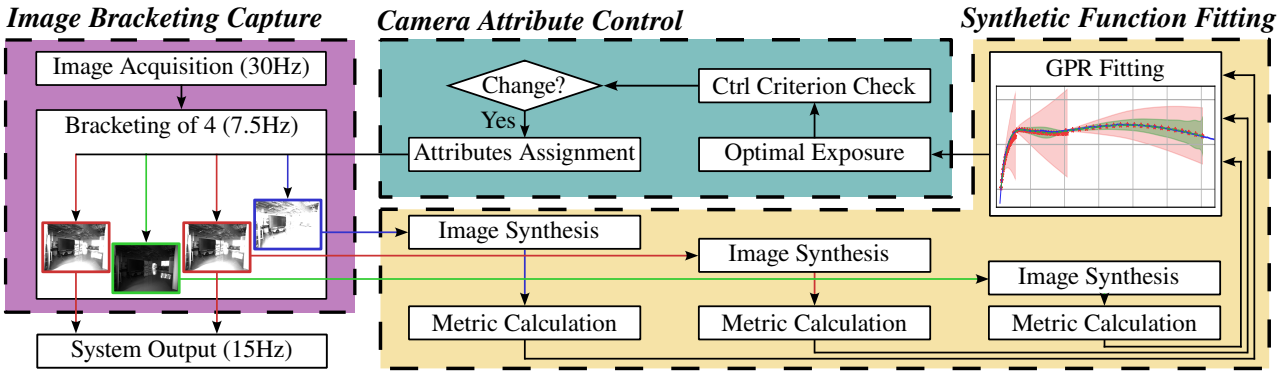


Fig. 1: Our system overview. By designing specific image bracketing patterns (Section IV-D), our system captures images with three different exposures and can explore the scene with a greater dynamic range. According to the camera imaging principle (Section IV-A), synthetic images are generated (Section IV-B), and their image quality metrics are evaluated (Section IV-C). Afterwards, the relationship between image quality metric and exposure is revealed, and GPR fitting technique (Section IV-E) estimates the optimal attributes for the next image. Finally, camera attribute control (Section IV-F) is implemented, and our system outputs images for VO systems.

combined with image simulation. They introduced optical flow to estimate the degree of the camera's ego motion to divide the gain and exposure time with motion blur awareness. However, gradient-descent-based methods introduce large errors when approximating the nonlinear camera response function with first-order Taylor expansion. Therefore, these methods always need an iterative manner instead of directly giving the optimal solution, which brings a trade-off between the optimization rate and the change rate of the scene. Some metrics' derivatives, such as image local entropy, are difficult to calculate. Methods can only choose metrics with analytical derivatives or calculate with numerical approximation.

Tomasi *et al.* [15] used a neural network to predict the exposure for future scenes on a vehicle VO system. They designed a proxy measure on a dual-camera setup to select the best-associated image pairs and trained the network under a reinforcement learning scheme. However, the training process is still complicated, and the generalization ability for scenarios has yet to be verified.

Several methods generate synthetic images to search the optimal attributes directly. Due to the image-level simulation, these methods require more calculations but directly give optimal output within one image. Shim *et al.* [16] simulated seed images under different exposure levels with γ-correction techniques. Gradient magnitude evaluates imaging quality, and a fifth-order polynomial fits the optimal gamma value. Since the relationship between gamma value and exposure cannot be directly revealed, the auto-adjust exposure control system still works under a feedback control scheme. Kim *et al.* [17], [18] designed an image quality measure considering gradient, entropy, and signal-to-noise ratio. Synthetic images are generated with a seed image using the photometric principle. The optimal exposure time and gain are explored using the Bayesian optimization method simultaneously. The only disadvantage of this method is that the seed image parameters are manually selected. A fixed seed image cannot ensure the performance when the scene has a wide dynamic range. Meanwhile, the input rate of the VO system will also decrease because seed grabbing halts image acquisition.

In this paper, we also propose a camera attribute control method based on image synthesis. Instead of inserting query seed images, we designed special image bracketing patterns, which avoid problems of manual seed selection and acquisition frequency drop. The acquisition system periodically explores the scenario's lighting up and down to understand the scenario irradiance better. With the perception of a higher dynamic range, our method adaptively gives a more reliable optimal exposure without any iterations or feedback control.

III. Problem Formulation

We refer to the system framework of [18], and our system architecture is illustrated as Fig. 1. The camera attribute control problem is formulated as an optimization problem:

$$\bar{e} = \arg\max_{e} \mathcal{H}(e) = \arg\max_{e} \mathcal{M}(\mathcal{S}(\boldsymbol{I}_0, e_0, e)). \quad (1)$$

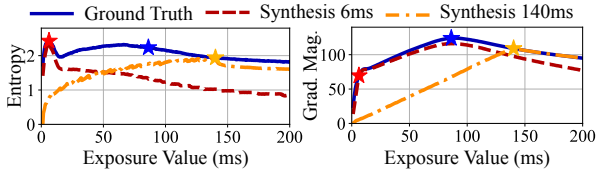
Image synthesizing function $\mathcal{S}(\cdot)$ transfers a captured image $\boldsymbol{I}_0$ with exposure e_0 to several synthetic images, and metric function $\mathcal{M}(\cdot)$ quantifies them. After synthesizing images from seeds, the function of the image quality metric w.r.t. exposure $\mathcal{H}(\cdot)$ is fitted. The optimal function value $\mathcal{H}(\bar{e})$ is ergodically searched, and the corresponding exposure $\bar{e}$ is selected for the next control.

IV. Methodology

A. Camera Response Function

We briefly review the imaging acquisition pipeline; more details can be found in [19], [20]. A camera sensor converts the photons collected during the exposure time of an image into a voltage signal. Analog gain is attached to amplify the perceived intensity of the scene. Afterward, a non-linear function remaps the signal to the intensity space and clips with the maximum intensity value. A round function finally turns the value into an integer intensity.

(a) The synthetic image from low exposure (dashed) appears mosaics due to quantization errors. The one from high exposure (dash-dot) reveals full textures, but overexposed areas cannot be recovered.



(b) Relationship between image quality and exposure using different synthetic images. The ground truth results are calculated by the images with different exposures in a static scene. Synthetic images are generated from original images of 6 ms (red star) and 140 ms (yellow star), respectively.

Fig. 2: Synthetic images and their relationships with two basic image quality metrics.

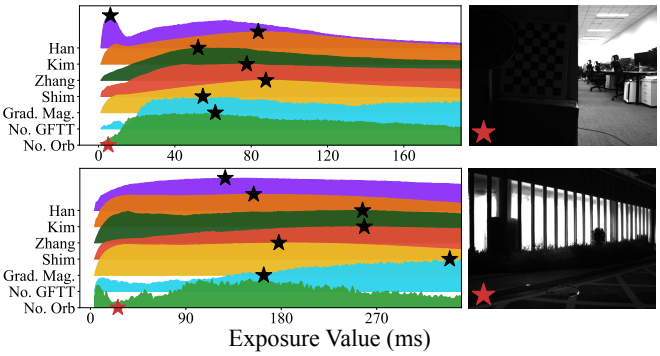


Fig. 3: Relationship between normalized image metrics and exposure value. The black stars represent the best exposures from different metrics, while the red represents the result from camera's built-in automatic exposure algorithm.

Camera response function (CRF) describes how the imaging system maps scene irradiance $\boldsymbol{E} \in \mathbb{R}^{W\times H}$ to an image $\boldsymbol{I} \in \mathbb{I}^{W\times H}$ (for 8-bit images $\mathbb{I} = \{i | i \in \mathbb{N} \text{ and } x \leq 255\}$),

$$\boldsymbol{I} = \mathcal{F}(e \cdot \boldsymbol{E}). \tag{2}$$

Exposure e is jointly determined by exposure time t and analog gain g (in dB) as $e = 10^{\frac{g}{20}} \times t$. CRF is not injective due to the round and truncation operations; thus it is mathematically not invertible. This can be solved by an inverse mapping from image intensity to logarithmic irradiance, known as the *inverse response function*,

$$\mathcal{G}(\boldsymbol{I}) = \ln e + \ln \boldsymbol{E}. \tag{3}$$

This discrete function $\mathcal{G}(\cdot)$ is invertible and can be obtained by the method of [19], returning a look-up table. We consider its inverse function on continuous space, nominated as $\mathcal{G}^{-1}(\cdot)$, fitted by the look-up table of $\mathcal{G}(\cdot)$ with a model of a tenth-order polynomial combined with Lasso regularization.

B. Photometric Image Synthesis

Similar to the work of [14], [18], we also consider image synthesis techniques to explore the relationship between image quality metric and exposure. When capturing a seed image $\boldsymbol{I}_0$ with the exposure e_0, the logarithmic scene irradiance is estimated by Equ. (3). A synthetic image $\boldsymbol{I}_1$ with the target exposure e_1 is then generated by $\mathcal{G}^{-1}(\cdot)$ after compensating for the difference in logarithmic exposure. The entire synthetic process can be represented as

$$\begin{aligned}\boldsymbol{I}_1 &= \mathcal{S}(\boldsymbol{I}_0, e_0, e_1) \\ &= \max\left(0, \min\left(255, \mathcal{G}^{-1}(\mathcal{G}(\boldsymbol{I}_0) - \ln e_0 + \ln e_1)\right)\right).\end{aligned} \tag{4}$$

The clip function ensures pixel value is between 0 and 255.

Image synthesis from different seed images is shown in Fig. 2a, and it is obvious to find the differences between the captured and the generated images. This discrepancy is generally caused by the quantization and truncation errors introduced by the imaging process. Quantization error from the round function mainly occurs when synthesizing images from lower exposure. The non-linear CRF varies the sensitivity of different pixel values to the corresponding irradiances. High pixel values are more sensitive and describe a narrower range of irradiance. After synthesizing from a low-exposure image, the pixel values cannot be sufficiently distinguished, and image mosaics appear. Truncation error occurs only when synthesizing images from higher exposure. The overexposed pixels (whose intensities are 255) cannot be recovered on the synthetic images. Due to these two errors, image metric calculations on synthetic images are not always reliable. As shown in Fig. 2b, gradient magnitude has strong applicability when synthesizing higher exposure within a certain range. The difference between the estimated and the ground truth values increases as the exposure change increases. This rule no longer exists when making lower exposure synthesis because of the loss of overexposed information. Image local entropy correlates with metric ground truth, but the relationship can hardly be quantitatively defined.

Instead of designing a specific metric, we use average gradient magnitude, the most general metric for image quality evaluation. In section IV-C, we will discuss image metrics for visual odometry tasks and show the adequacy and rationality of selecting this metric.

C. Image Quality Metric

Image quality metrics for camera attribute control are roughly divided into two categories. One effective and convincing way is to count feature quantity, which directly affects the front-end tracking performance of a VO system. However, feature extraction and matching are time-consuming and usually undifferentiable to obtain an optimization direction. Therefore, feature quantity is only used as a metric in a learning-based framework [15]. Another thought uses image statistical properties related to feature extraction performance. Gradient magnitude shows the pixel-wise intensity change rate. The gradient magnitude of pixel

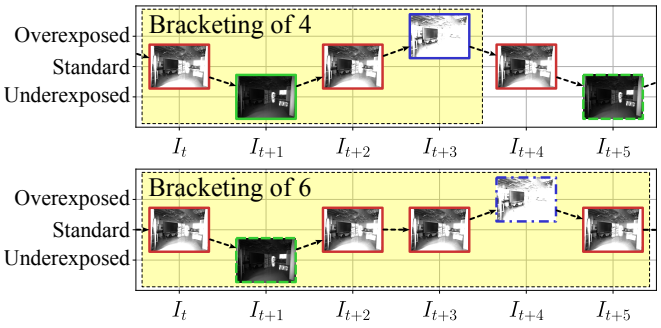


Fig. 4: Two different image bracketing patterns. Suppose the camera acquisition frequency is 30 Hz. For bracketing of 4 (images), the system output and theoretical maximum control frequency are both 15 Hz. The frequencies turn to 20 Hz and 10 Hz when using a bracketing of 6. More bracketing patterns can be designed to balance the output and control frequencies.

I_i is defined as

$$\mathcal{M}_{gm}(I_i) = ||\nabla I_i||_2,$$

and is calculated by Sobel operator in practice. Local entropy indicates the statistical randomness of an image patch

$$\mathcal{M}_{le}(I_i) = -\sum_{j=0}^{255} p(j)\log_2 p(j),$$

where $p(j)$ represents the probability that intensity j appears within the image patch around pixel i. Local entropy directly identifies image overexposed and underexposed regions.

As shown in Fig. 3, we evaluate different image quality metrics with the quantity of two specific features (ORB [21] and GFTT [22]). All the metrics show suitable effects and obtain sufficient features. We cannot assert which metric is better because the scores are affected by several factors, such as scene, camera field of view, and image resolution. In this paper, we choose average gradient magnitude

$$\mathcal{M}(\boldsymbol{I}) = \frac{1}{N}\sum_{i=0}^{N}\mathcal{M}_{gm}(I_i), \tag{5}$$

for its computational efficiency and numerical smoothness, where N represents the number of pixels in this image. We downsample the captured or synthesized images before metric calculation in order to filter out noise from the analog gain and synthesis process. This operation also removes high-frequency textures that bring unstable feature points for VO systems and preserves stable scene structure information.

D. Image Bracketing Capture

Image bracketing is a powerful photographic technique in HDR scenes. Our camera captures images with three exposures of the same scene: a standard image, a darker (underexposed) image, and a brighter (overexposed) image. The standard images are published for subsequent VO systems, while the underexposed and overexposed images are only used for scene irradiance exploration. The imaging system obtains a wider dynamic range, and the scene irradiance is sufficiently accessed. Considering image freshness, the attribute control module is triggered only for non-standard images. Two image bracketing patterns are implemented (Fig. 4), balancing the system output and control frequency.

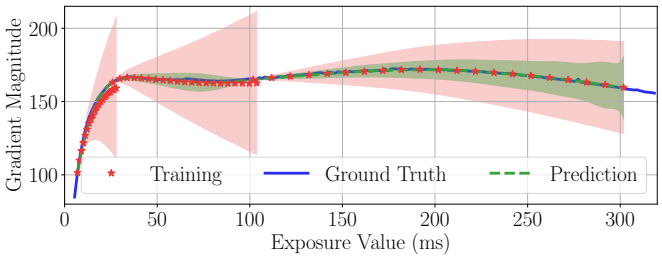


Fig. 5: Demonstration of the training inputs and prediction outputs of our GPR. The training inputs with uncertainty are collected from three different exposure levels in a bracket. The prediction outputs are close to the ground truth and can resist the errors from image synthesis.

Once the optimal exposure for the current standard image is calculated (details in Section IV-F), the underexposed and overexposed exposures are passively determined. A empirically set ratio μ controls the difference between adjacent image levels, where $e_{\text{std}} = \mu \cdot e_{\text{ue}}$ and $e_{\text{oe}} = \mu \cdot e_{\text{std}}$.

E. Synthetic Function Fitting

Gaussian process regression (GPR) is chosen to fit function $\mathcal{H}(\cdot)$, which reveals the function with its uncertainty. Suppose we have the training data $\{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\sigma}^2\}$, where $\boldsymbol{x} = \boldsymbol{e} \in \mathbb{R}^N$ and $\boldsymbol{y} = \mathcal{H}(\boldsymbol{e}) \in \mathbb{R}^N$ are exposures and their corresponding image metric values, and $\boldsymbol{\sigma}^2 \in \mathbb{R}_+^N$ are the uncertainty on $\mathcal{H}(\boldsymbol{e})$. The predicted metric value y_* and its uncertainty σ_*^2 for x_* can be formulated as

$$y_* = \boldsymbol{K}_*^\top(\boldsymbol{K} + \text{diag}(\boldsymbol{\sigma}^2))^{-1}\boldsymbol{y},$$
$$\sigma_*^2 = k_{**} - \boldsymbol{K}_*^\top(\boldsymbol{K} + \text{diag}(\boldsymbol{\sigma}^2))^{-1}\boldsymbol{K}_*,$$

where $\text{diag}(\cdot)$ convert a vector to a diagnoal matrix. The kernel function $\mathcal{K}(\cdot,\cdot)$ is defined as a radial basis function,

$$\mathcal{K}(x_1, x_2) = s^2 \exp\left(-\frac{||x_1 - x_2||^2}{2l^2}\right),$$

where s and l are offline pretrained hyperparameters. Matrices $\boldsymbol{K}$, $\boldsymbol{K}_*$ and scalar k_{**} are derived by the kernel function as $\boldsymbol{K} = \mathcal{K}(\boldsymbol{x}, \boldsymbol{x}) \in \mathbb{R}_+^{N\times N}$, $\boldsymbol{K}_* = \mathcal{K}(\boldsymbol{x}, x_*) \in \mathbb{R}_+^N$ and $k_{**} = \mathcal{K}(x_*, x_*) \in \mathbb{R}_+$.

In the work of [18], the training data is generated by the Bayesian optimization, searching for the optimal query point and updating the GPR model iteratively, with a constantly defined uncertainty. Unlike their work, we uniformly sample synthetic images within a valid range. According to IV-B, image synthesis to lower exposures is unreliable, and higher exposure synthesis is valid within a certain range. Therefore, for an image seed with exposure e_{seed}, we uniformly synthesize n images between e_{seed} and $\mu \cdot e_{\text{seed}}$. The expected

exposure for i-th synthetic image is

$$e_{\text{syn},i} = e_{\text{seed}} + \frac{i}{n} \cdot (\mu - 1) \cdot e_{\text{seed}}, \tag{6}$$

where μ is the ratio defined in IV-D. We correlate the relationship between uncertainty and the exposure distance between samples and the seed. Since seed images have no synthetic error, its uncertainty $\sigma^2_{e_{\text{seed}}}$ remains 0. The maximum uncertainty occurs when exposure is $\mu \cdot e_{\text{seed}}$ and is defined as $\sigma^2_{e_{\mu \cdot \text{seed}}} = \alpha^2 \mathcal{H}(e_{\text{seed}})$, where α is a ratio between 0 and 1. The metric uncertainty for i-th synthetic image is

$$\sigma^2_{syn,i} = \frac{e_{\text{syn},i} - e_{\text{seed}}}{(\mu - 1) \cdot e_{\text{seed}}} \cdot \sigma^2_{\mu \cdot \text{seed}}. \tag{7}$$

The training input and prediction output are shown in Fig. 5. When a control is required, the latest images in each level are selected as seeds. Training data are separately generated from seeds and afterward aggregated to train the GPR model.

F. Attribute Control Strategy

1) Camera Attribute Assignment: Camera exposure e is achieved by the combination of exposure time t and gain g, and it needs to be assigned to these two attributes when using a camera software interface in practice. The trade-off between exposure time and gain involves two aspects. A long exposure time may cause blurry images, especially for VO tasks where cameras move and rotate rapidly. Gain amplifies image signal as well as image dark noise, and the relationship between the noise level and the gain is in second order [20]. Han [14] considered camera motion blur and introduced optical flow to estimate the degree of motion. They use the median of optical flow motion as an indicator to limit the maximum exposure time. Although the subsequent VO system will probably introduce optical flow calculation, it is still burdensome for camera attribute control tasks and not beneficial for system independence.

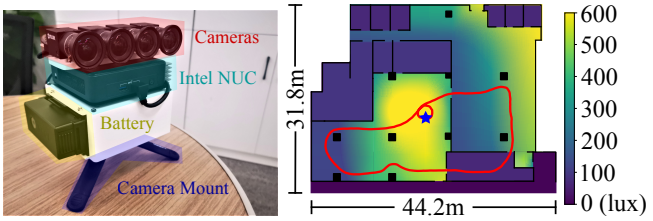
We propose a progressive strategy for attribute assignment by several intermediate attribute levels. We rate the target exposure into one of these intervals and afterward assign the attribute with exposure time priority. For instance, considering three levels of [5 ms, 5 dB] (8.89 ms), [10 ms, 10 dB] (31.62 ms), and [20 ms, 20 dB] (200 ms). A target exposure of 10 ms is assigned between [5 ms, 5 dB] and [10 ms, 10 dB], and the expected attribute is [5.62 ms, 5 dB].

2) Control Activation Condition: Considering the flickering effect and reducing the control frequency, a global illumination difference criterion [18] judges whether to modify attributes for the next image. We also extend a ratio criterion to avoid control requests when the absolute exposure difference is small.

V. Experiments

A. Experimental Setup

A handheld platform (Fig. 6a) is designed for experimental evaluation. Four FLIR BFS-U3-31S4C cameras are deployed to compare attribute control methods, which are implemented on an Intel NUC11TNKi7 and powered by a portable battery.



(a) Our Experimental platform. (b) Scene luminance and test trajectory.

Fig. 6: Our experimental setup. The scene luminance is measured by a photometer. The evaluation trajectory (marked in red) is connected end to end, with endpoint marked in blue.

Our system (**Ours**) uses an image bracketing pattern of 4. The camera capturing frequency is 30 Hz, and the system output frequency is 15 Hz. The ratio μ is 4, and the number of synthetic image samples n is 20, balancing the GPR fitting performance and real-time computing. The uncertainty level α is empirically set to 0.3. For real-time efficiency, our system uses Intel AVX to accelerate the image synthesis, and it takes 0.636 ms to synthesize an image of 512×384.

Three attribute control methods are chosen as baseline methods. FLIR camera comes with a **Built-in** exposure control algorithm, a feedback control system that optimizes image average pixel intensity. The methods of **Shim** [16] and **Kim** [18] are chosen and implemented via the FLIR Spinnaker interface. For a fair comparison, all the baseline methods capture images at 30 Hz but publish to VO systems at a frequency of 15 Hz. Control delay always occurs for industrial cameras, and we find that FLIR's external control interface always suffers from a 2-frame delay in practice. **Built-in** is embedded in the camera and can reach a control frequency of 30 Hz, the same as the image acquisition frequency. **Shim** uses open-loop feedback control. Although the attributes are adjusted at 30 Hz, its effective control frequency is only about 10 Hz. **Kim** needs an additional seed image sampling, which means 6 frames are required (maximum 5 Hz) for a complete control cycle. Although no seed image is required, **Ours** needs a full bracketing to understand the scene. For bracketing of 4, our system also needs 6 frames to achieve control.

B. Visual Odometry Evaluation

An image sequence is captured by our platform in the complex lighting scene of our laboratory. The trajectory and lighting conditions are shown in Fig. 6b.

We evaluate the number of detected and tracked ORB features of the attribute control methods, shown in Fig. 7. **Ours** maintains high feature detection and tracking rates for most sequence frames. Images collected by our system are slightly overexposed because we use the original gradient magnitude, where the overexposure phenomenon is not mitigated by local entropy mask or statistic features. In practice, properly overexposed images benefit the extraction and identification of feature points [23].

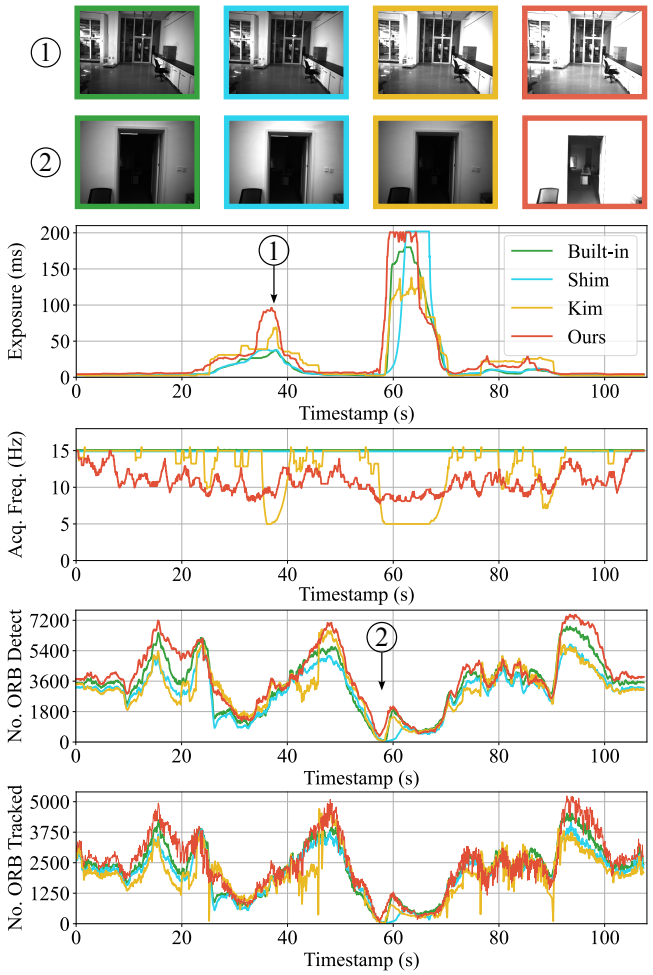


Fig. 7: Evaluations of the test sequence, including exposure, acquisition frequency, and the ORB feature detection and tracking quantity. The first case shows that the images of Kim's and ours are brighter in dark scenes, and more features are detected. The second one shows that our method perceives the existence of rich information in the dark regions in advance, benefiting from our image bracketing mechanism. In addition, our acquisition frequency cannot be maintained at 15 Hz because the camera interface cannot perfectly adapt to our system's requirements, and additional time overhead is brought when adjusting the bracketing parameters.

We also evaluate the VO performance by piping the images captured from each method into the ORB-SLAM2 [24]. Since the trajectory is connected end to end, we evaluate the endpoint's absolute trajectory error (ATE). As shown in Fig. 8, **Ours** has the best VO performance, with an ATE of 0.854 m, while **Built-in** (1.085 m, 26.9% larger) and **Shim** (1.092 m, 27.8% larger) have similar results. **Kim** has a high-quality trajectory until the system fails due to drastic lighting changes and frequency drop caused by seed query.

C. Dramatic Illumination Change

We capture a static sequence containing light-on and light-off events to test system control response speed under

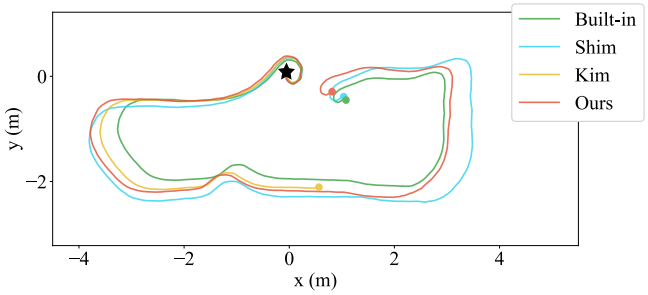


Fig. 8: Visual odometry trajectories after scale normalization. The star represents the start position of this trajectory while the dots represent the end positions from different methods.

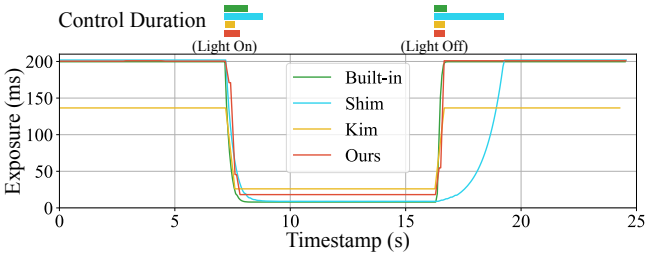


Fig. 9: Exposure control for light-on and light-off events. One-step control methods (**Kim** and **Ours**) respond rapidly. Feedback control schemes (**Built-in** and **Shim**) need longer control durations. Images are oversaturated (overexposed and underexposed) during the initial period of control, which may greatly reduce the detection and tracking rate of features.

dramatic illumination changes. The scene illumination varies greatly, while the light-on and light-off luminances are 500 lx and 3 lx, respectively. As shown in Fig. 9, both **Kim** and **Ours** can respond rapidly to drastic light changes, with the control durations of 429 ms and 642 ms, respectively. This benefits from the direct calculation of target exposure without iterative optimization. **Built-in** and **Shim** are feedback control systems, requiring longer control durations. **Built-in** also gives a quick response (1000 ms for light-on and 600 ms for light-off) for its rapid internal control interface. **Shim** method requires a moderate control duration of 1750 ms for the light-on event. However, the light-off event takes 3250 ms because they use different convergence rates for different directions.

VI. Conclusions and Future Work

In this paper, we propose an image acquisition system to alleviate the challenges of visual odometry tasks under complex illumination scenes. Image bracketing patterns and camera attribute control module assist our system in exploring the scene luminance sufficiently. Gaussian process regression and image synthesis techniques quickly traverse the optimal exposure and ensure a quick response to scene lighting change. Experimental evaluations show the effectiveness and robustness of our system for VO tasks.

In our future work, we plan to design camera hardware and drivers to fix the drop in acquisition frame rate (Fig. 7). We also plan to improve our system on fisheye cameras and afterward extend it to a complete visual SLAM pipeline.

REFERENCES

[1] S. Park, T. Schöps, and M. Pollefeys, "Illumination change robustness in direct visual slam," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 4523–4530.

[2] M. Bujanca, X. Shi, M. Spear, P. Zhao, B. Lennox, and M. Luján, "Robust slam systems: Are we there yet?" in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 5320–5327.

[3] K. Xu, Y. Hao, S. Yuan, C. Wang, and L. Xie, "Airvo: An illumination-robust point-line visual odometry," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023.

[4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[5] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The euroc micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.

[6] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers, "The tum vi benchmark for evaluating visual-inertial odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1680–1687.

[7] A. R. Vidal, H. Rebecq, T. Horstschaefer, and D. Scaramuzza, "Ultimate slam? combining events, images, and imu for robust visual slam in hdr and high-speed scenarios," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 994–1001, 2018.

[8] K. Ye, L. Gao, and B. Guan, "Visual odometry in hdr environments by using spatially varying exposure camera," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 5995–6000.

[9] R. Gomez-Ojeda, Z. Zhang, J. Gonzalez-Jimenez, and D. Scaramuzza, "Learning-based image enhancement for visual odometry in challenging hdr environments," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 805–811.

[10] E. Jung, N. Yang, and D. Cremers, "Multi-frame gan: Image enhancement for stereo visual odometry in low light," in *Conference on Robot Learning*. PMLR, 2020, pp. 651–660.

[11] J. Kim, M.-H. Jeon, Y. Cho, and A. Kim, "Dark synthetic vision: Lightweight active vision to navigate in the dark," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 143–150, 2020.

[12] Z. Zhang, C. Forster, and D. Scaramuzza, "Active exposure control for robust visual odometry in hdr environments," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 3894–3901.

[13] Y. Wang, H. Chen, S. Zhang, and W. Lu, "Automated camera-exposure control for robust localization in varying illumination environments," *Autonomous Robots*, vol. 46, no. 4, pp. 515–534, 2022.

[14] B. Han, Y. Lin, Y. Dong, H. Wang, T. Zhang, and C. Liang, "Camera attributes control for visual odometry with motion blur awareness," *IEEE/ASME Transactions on Mechatronics*, 2023.

[15] J. Tomasi, B. Wagstaff, S. L. Waslander, and J. Kelly, "Learned camera gain and exposure control for improved visual feature detection and matching," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 2028–2035, 2021.

[16] I. Shim, J.-Y. Lee, and I. S. Kweon, "Auto-adjusting camera exposure for outdoor robotics using gradient information," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 1011–1017.

[17] J. Kim, Y. Cho, and A. Kim, "Exposure control using bayesian optimization based on entropy weighted image gradient," in *2018 IEEE International conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 857–864.

[18] ——, "Proactive camera attribute control using bayesian optimization for illumination-resilient visual navigation," *IEEE Transactions on Robotics*, vol. 36, no. 4, pp. 1256–1271, 2020.

[19] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, 2008, pp. 1–10.

[20] Y. Wang, H. Huang, Q. Xu, J. Liu, Y. Liu, and J. Wang, "Practical deep raw image denoising on mobile devices," in *European Conference on Computer Vision*. Springer, 2020, pp. 1–16.

[21] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International conference on computer vision*. Ieee, 2011, pp. 2564–2571.

[22] J. Shi *et al.*, "Good features to track," in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. IEEE, 1994, pp. 593–600.

[23] N. Yang, R. Wang, X. Gao, and D. Cremers, "Challenges in monocular visual odometry: Photometric calibration, motion bias, and rolling shutter effect," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 2878–2885, 2018.

[24] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE transactions on robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.