Active Exposure Control for Robust Visual Odometry in HDR Environments

Zichao Zhang, Christian Forster, Davide Scaramuzza

***Abstract*— We propose an active exposure control method to improve the robustness of visual odometry in HDR (high dynamic range) environments. Our method evaluates the proper exposure time by maximizing a robust gradient-based image quality metric. The optimization is achieved by exploiting the photometric response function of the camera. Our exposure control method is evaluated in different real world environments and outperforms the built-in auto-exposure function of the camera. To validate the benefit of our approach, we adapt a state-of-the-art visual odometry pipeline (SVO) to work with varying exposure time and demonstrate improved performance using our exposure control method in challenging HDR environments.**

SUPPLEMENTARY MATERIALS

A video demonstrating the improvement on different visual odometry algorithms is available at `https://youtu.be/TKJ8vknIXbM`.

I. INTRODUCTION

Recently, VO (visual odometry) algorithms have reached a high maturity and there is an increasing number of applications in various fields, such as VR/AR. Although many impressive results have been presented, one of the remaining challenges is robustness in HDR environments. The difficulty in such environments comes from the limitations of both the sensor and the algorithm. For conventional cameras, the dynamic range is narrow compared to real world environments. Without proper exposure control, images can be easily overexposed or underexposed, and very little information can be extracted from such images. In order to overcome the problem of the narrow dynamic range, many cameras automatically adjust the exposure time. The change of exposure time, however, breaks the brightness constancy assumption across consecutive frames, which is the underlying assumption of many VO algorithms. Therefore, to work in HDR environments, a VO algorithm should be *active*, instead of passive. An active VO algorithnm, on the one hand, must actively adjust the exposure time of the camera to maximize the information for VO; on the other hand, the effect of the varying exposure time needs to be explicitly compensated .

While the topic of exposure control has been studied extensively, little work has been done to optimize the exposure time for VO applications. Moreover, most exposure control methods rely on heavily engineered parameters, because of the lack of a quantitative knowledge on how the change of the exposure time affects the image. Regarding exposure compensation, a widely used technique is to model the brightness change with an affine transformation. Alternatively, researchers have recently exploited the photometric response function of the camera for exposure compensation [1], [2]. While both methods are shown to work, to the best of our knowledge, there is no comparison study of them yet in the existing literature. It would be interesting to know, from a practical perspective, which compensation method should be used when building VO applications.

In this paper, we first propose an active exposure control method to maximize the gradient information in the image. This is inspired by the observation that most vision algorithms, including VO, actually extract information from gradient-rich areas. For instance, corners are essentially points where the gradient is large in two orthogonal directions [3]; direct VO algorithms also make use of the pixels with high gradients [2], [4]. Therefore, we propose a gradient-based image quality metric and show that it is robust in HDR environments by an extensive evaluation in different scenarios. Moreover, we use the photometric response function of the camera to design our exposure control scheme. By exploiting the photometric response function, we are able to evaluate the derivative of our metric with respect to the exposure time. Such information enables us to apply mathematically grounded methods, such as gradient descent, in exposure control. Second, we introduce our adaptations of exposure compensation to a state-of-the-art VO algorithm, namely SVO (Semi-direct Visual Odometry [5]). We formulate these adaptations in an algorithm-agnostic manner, so that they can be easily generalized to other VO algorithms. In addition, an experimental comparison of the aforementioned exposure compensation methods is presented. Finally, we demonstrate in several real-world experiments that, with the proposed exposure control method, our VO algorithm is able to operate in HDR environments.

A. Related Work

Many existing exposure control approaches use heuristics based on image statistics, such as the mean intensity value and the intensity histogram. A system for configuring the camera parameters was presented in [6]. The exposure time was selected according to the intensity histogram of the image. Their method was successfully used in practice during the RoboCup competitions [7]. More recently, Torres *et al.* [8] used a set of indicators from the intensity histogram and the cumulative histogram to capture the different aspects of the image quality, and a camera exposure control method

The authors are with the Robotics and Perception Group, University of Zurich, Switzerland—`http://rpg.ifi.uzh.ch`. This work was supported by the China Scholarship Council, the DARPA FLA Program, the NCCR Robotics through the Swiss National Science Foundation, and the SNSF-ERC Starting Grant.

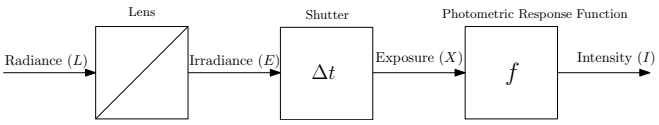


Fig. 1: Image Acquisition Process

was designed based on these indicators.

By contrast, other works explicitly explore the information in the image. Lu *et al.* [9] characterized the image quality using Shannon's entropy. They showed experimentally that the entropy of the image was related to the performance of the object recognition algorithm. Therefore, the exposure control was achieved by searching for the highest entropy in the parameter space of the camera. Closely related to our work is [10], which used the gradient information within an image to select the proper exposure time. The authors defined an information metric based on the gradient magnitude at each pixel. The exposure change was simulated by applying different gamma corrections to the original image to find the gamma value that maximizes the gradient information. Then, the exposure time was adjusted based on the gamma value. Our work differs from [10] in two aspects. First, we use a different gradient-based metric, which we demonstrate to be more robust. Second, our control scheme also exploits the photometric response function of the camera.

Different methods have been proposed for exposure compensation. Jin *et al.* [11] used an affine transformation to model the illumination change in the feature tracking problem and showed success tracking under significant illumination changes. Kim *et al.* [12] jointly estimated the feature displacements and the camera response function and used the estimated response function to improve the performance of feature tracking. More recently, Engel *et al.* [2] used an affine brightness transfer function to compensate for the variation of the exposure time and applied it to VO. In addition, they also proposed to use the photometric response function of the camera for exposure compensation if the exposure time of the camera is known. Similarly, Li *et al.* [1] exploited the camera response function to account for the brightness change caused by the auto-exposure of the camera and applied it to a tracking and mapping system.

After introducing the photometric response function in Section II, we propose our gradient-based image quality metric in Section III. Based on the photometric response function and the image quality metric, our exposure control method is described in Section IV. Then, in Section V, we describe our adaptations of exposure compensation to a VO algorithm and compare the two commonly used exposure compensation techniques mentioned above experimentally. Finally, we validate our exposure control algorithm and demonstrate robust VO in HDR environments in Section VI.

II. Photometric Response Function

In this work, we use the photometric response function proposed in [13]. For completeness, we briefly introduce the function in the following.

The image formation process is illustrated in Fig. 1. For each pixel, the *irradiance* E describes the amount of energy

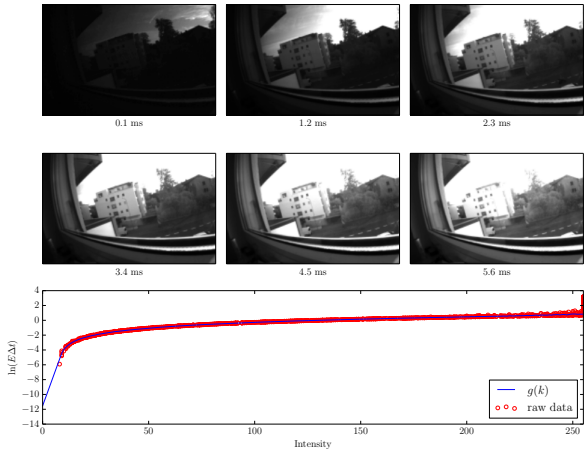


Fig. 2: The top two rows are images captured under different exposure times, used as the input to the calibration. The third row shows the recovered inverse response function.

that hits the pixel per time unit, and the *exposure* X is the total amount of energy received by the pixel during the exposure time Δt. The *photometric response function* f maps the exposure X to the intensity I in the image:

$$I = f(X) = f(E\Delta t). \tag{1}$$

Note that $f(\cdot)$ is invertible because the intensity should increase monotonically with the exposure. Then, for convenience, we can define the *inverse response function*

$$g = \ln f^{-1}, \tag{2}$$

and (1) can be written as

$$\tilde{g}(I) \quad = \quad \ln E + \ln \Delta t. \tag{3}$$

Obviously, for a digital image, where the possible intensities are a range of discrete values $\{0, 1, \ldots, Z_{\max}\}$, $\tilde{g}$ can only take values $g(k), k = 0, 1, \ldots, Z_{\max}$. These values can be determined by analyzing the images of a static scene captured under different exposure times. For the details of the photometric calibration process, we refer the reader to [13]. A sample calibration sequence and the recovered inverse response function g are illustrated in Fig. 2. After recovering g, we estimate a tenth order polynomial to fit the discrete values in (3) and use the polynomial to calculate the derivative g'.

In the next section, the image quality metric used in our exposure control method is introduced.

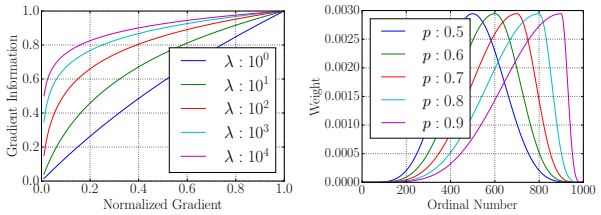
III. Image Quality Metrics

The metrics for image quality are highly application-dependent. Regarding VO applications, the gradient information is of great importance for both feature-based and direct methods. In this section, we first introduce several gradient-based metrics and then compare them on real world data.

A. Gradient-Based Metrics

Given an image, denoted as I, captured with an exposure time Δt, the magnitude of the gradient at a pixel $\mathbf{u}$ is

$$G(\mathbf{I}, \mathbf{u}, \Delta t) \quad = \quad \|\nabla \mathbf{I}(\mathbf{u}, \Delta t)\|^2, \tag{4}$$



(a) Information in M_{shim} (b) M_{perc} weights ($k = 5$).

Fig. 3: The mapping function of M_{shim} and the weights in M_{softperc}

where $\nabla \mathrm{I}(\cdot) = [\frac{\partial \mathrm{I}}{\partial x}, \frac{\partial \mathrm{I}}{\partial y}]^\top$. In the rest of this section, we drop the notation of I in (4) for simplicity.

A straightforward metric is the sum of (4) on all the pixels in the image:

$$M_{\text{sum}} = \sum_{\mathbf{u}_i \in \mathbf{I}} G(\mathbf{u}_i). \tag{5}$$

Alternatively, Shim *et al.* [10] defined the *gradient information* of a pixel $\mathbf{u}_i$ as

$$m_{\mathbf{u}_i} = \begin{cases} \frac{1}{N} \log(\lambda(\tilde{G}(\mathbf{u}_i) - \sigma) + 1), & G(\mathbf{u}_i) \geq \sigma \\ 0, & G(\mathbf{u}_i) < \sigma \end{cases}, \tag{6}$$

where $\tilde{G}$ is the gradient magnitude normalized to the range of $[0, 1]$, $N = \log(\lambda(1 - \sigma) + 1)$ is a normalization factor to bound the gradient information to the range of $[0, 1]$, σ is an activation threshold, and λ determines whether strong or weak intensity variations are emphasized. Then the total gradient information in an image is

$$M_{\text{shim}} = \sum_{\mathbf{u}_i \in \mathbf{I}} m_{\mathbf{u}_i}. \tag{7}$$

M_{shim} can be interpreted as a weighted sum of the gradient magnitudes from all the pixels. The mapping from the normalized gradient magnitude $\tilde{G}$ to the gradient information (6) is plotted in Fig. 3a for different λs. For both M_{sum} and M_{shim}, the main problem is that the squared sum is not a robust estimator of the scale of the gradient magnitudes (see Section III-B).

Instead, we consider using a certain percentile of all the gradient magnitudes as a robust estimator:

$$M_{\text{perc}}(p) = \text{percentile}(\{G(\mathbf{u}_i)\}_{\mathbf{u}_i \in \mathbf{I}}, p), \tag{8}$$

where p indicates the percentage of the pixels whose gradient magnitudes are smaller than M_{perc}. For example, M_{perc} is the median of all the gradient magnitudes when $p = 0.5$.

Lastly, we define another gradient-based metric, which is called *soft percentile* in this paper. We first sort the gradient magnitudes of all the pixels $\{G(\mathbf{u}_i)\}_{\mathbf{u}_i \in \mathrm{I}}$ in an ascending order. The sorted gradient magnitudes are denoted as $\{G_{i\text{th}}\}_{i \in [0,S]}$, where S is the total number of pixels in the image. Then we calculate the soft percentile metric as a weighted sum of the sorted gradient magnitudes:

$$M_{\text{softperc}}(p) = \sum_{i \in [0,S]} W_{i\text{th}}(p) \cdot G_{i\text{th}}. \tag{9}$$

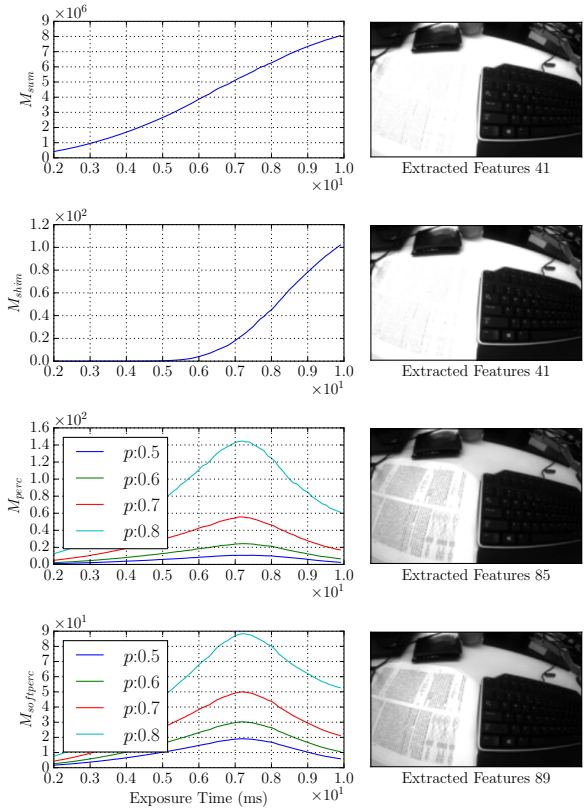


Fig. 4: A HDR scene. The left column illustrates how different metrics change with the exposure time. The right column shows the best image in terms of each metric, respectively.

The weights $\{W_{i\text{th}}(p)\}_{i \in [0,S]}$ are

$$W_{i\text{th}} = \begin{cases} \frac{1}{N} \sin(\frac{\pi}{2\lfloor p \cdot S \rfloor} i)^k, & i \leq \lfloor p \cdot S \rfloor \\ \frac{1}{N} \sin(\frac{\pi}{2} - \frac{\pi}{2} \frac{i - \lfloor p \cdot S \rfloor}{S - \lfloor p \cdot S \rfloor})^k, & i > \lfloor p \cdot S \rfloor \end{cases}, \tag{10}$$

where $\lfloor \cdot \rfloor$ rounds a number down to the closest integer, and N normalizes the sum of $\{W_{i\text{th}}(p)\}_{i \in [0,S]}$ to 1.

The weight function (10) is plotted in Fig. 3b for different values of p. Intuitively, the soft percentile approximates a certain percentile with a weighted sum of the gradient magnitudes. The larger the k, the closer M_{softperc} is to M_{perc}. The advantage of the soft percentile metric over the percentile metric is that it changes smoothly with the exposure time, which we will see next.

B. Evaluation

In order to understand the difference of the aforementioned metrics, we evaluate them on 18 real world datasets. Each of the datasets consists of a sequence of images of the same scene, captured with different exposure time settings. In the evaluation, we compute the gradient-based metrics for all the images and observe how different metrics change with the exposure time. For M_{softperc}, $k = 5$ is used. For both M_{perc} and M_{softperc}, the best image is chosen as the one that corresponds to the maximum value when $p = 0.5$ (as we will see, the best image does not change much with different values of p). To quantitatively measure the image quality, we compute the number of FAST features [14] that can be extracted from the best images.

TABLE I: Number of FAST features in the best image. The percentile-based metrics performs better in 13 out of 18 datasets.

Dataset	M_{sum}	M_{shim}	M_{perc}	M_{softperc}
office window1	272	272	**288**	**288**
office window2	23	23	**33**	**33**
building	**66**	61	62	62
office desk	336	336	**403**	391
office ceiling	**490**	429	456	456
keyboard	60	59	**84**	**84**
light	32	7	**34**	**34**
office door	55	53	**59**	**59**
home window1	67	**69**	67	67
home window2	**49**	46	45	46
corridor	**26**	**26**	**26**	**26**
clutter	**51**	50	**51**	49
shelf	78	**81**	80	78
lounge	81	50	**82**	**82**
garage	**66**	**66**	**66**	**66**
shady building	**100**	74	**100**	**100**
sunny building	**71**	**71**	**71**	**71**
grass	**81**	**81**	**81**	**81**

The results are listed in Table. I. It can be observed that in 13 out of 18 datasets, the best images in terms of M_{perc} have the most features. In addition, in the datasets where another metric performs better, the numbers of the features in the best images of M_{perc} are actually very close to those of the best metric (e.g., in *home window1*, 67 features compared to 69 features of M_{shim}). In contrast, in some datasets, much less features can be extracted from the best images of other metrics (e.g., *keyboard* and *lounge* datasets). The performance of M_{softperc} is quite close to M_{perc}.

To give an intuition of the difference among the metrics, we show the results of the *keyboard* dataset in Fig. 4. The scene mostly consists of two areas with very different brightness, a black keyboard and a piece of white paper with text. It can be observed that both M_{sum} and M_{shim} increase with the exposure, and the best images according to these metrics are obviously overexposed in the bright area (i.e., the piece of paper). In contrast, the best images in terms of the percentile based metrics, M_{perc} and M_{softperc}, preserve the details in the bright area well.

There are two observations worthnoting regarding the percentile based metrics. While M_{perc} and M_{softperc} have quite similar performance in our evaluation, if the plots of M_{perc} and M_{softperc} in Fig. 4 are closely compared, it can be seen that the curves corresponding to M_{softperc} are smoother. In addition, while the curves of different p values have similar maxima, the one corresponding to a higher p usually has larger derivative with respect to the exposure time. This is because, in an image, there are usually a large number of pixels with low gradient magnitudes under all exposure times (e.g., smooth area), which will make the percentiles with small p values change less significant. Both the smoothness and the derivatives are important for our optimization-based exposure control algorithms, which will be discussed in more details in Section IV. Based on the above observations, we will use M_{softperc} and $p = 0.7 \sim 0.8$ in the rest of the work.

To summarize, in our evaluation, the percentile based metrics M_{perc} and M_{softperc} are more robust than M_{sum} and M_{shim}, and M_{softperc} with a large p has a more desirable behavior. In the next section, we will describe our exposure control method.

IV. Exposure Control

With the photometric response function in Section II, we are able to predict how the image changes with the exposure time and, furthermore, we know how the metrics in Section III-A change accordingly. Such information allows us to use standard optimization methods, such as gradient descent, for exposure control. Following this idea, in this section, we first derive the derivative of the soft percentile metric (9) with respect to the exposure time and then describe our exposure control method.

A. Derivative of the Gradient Magnitude

Because our metric is based on the image gradient magnitude, the first step is to calculate the derivative of the squared gradient magnitude $G(\cdot)$ with respect to the exposure time Δt. Taking the derivative of the right-hand side of (4), $\frac{\partial G(\cdot)}{\partial \Delta t}$ becomes

$$2\nabla \mathrm{I}(\mathbf{u}, \Delta t)^\top \frac{\partial}{\partial \Delta t}[\nabla \mathrm{I}(\mathbf{u}, \Delta t)]. \tag{11}$$

The first term of (11) is simply the gradient of the image, and the second term can be transformed by applying the Schwarz's theorem:

$$\frac{\partial}{\partial \Delta t}[\nabla \mathrm{I}(\mathbf{u}, \Delta t)] \quad = \quad \nabla[\frac{\partial}{\partial \Delta t}\mathrm{I}(\mathbf{u}, \Delta t)]. \tag{12}$$

Note that the derivative inside the right-hand side of (12) is actually the derivative of the photometric response function (1). Thus, for a pixel with the intensity I, the derivative can be calculated as

$$\frac{\partial I}{\partial \Delta t} \overset{(1)}{=} f'[f^{-1}(I)]E(\mathbf{u}) = \frac{E(\mathbf{u})}{[f^{-1}]'(I)} \overset{(2)}{=} \frac{1}{g'(I)\Delta t}, \tag{13}$$

where $E(\mathbf{u})$ is the exposure corresponding to the pixel. Finally, inserting (13) into (12) and then (12) into (11), the derivative of the gradient magnitude becomes

$$\frac{\partial G(\cdot)}{\partial \Delta t} \quad = \quad 2[\nabla \mathrm{I}(\cdot)]^\top \nabla \left[\frac{1}{g'(\mathrm{I}(\cdot))\Delta t}\right]. \tag{14}$$

Note that $g'(\mathrm{I}(\cdot))$ means applying g' to all pixels of I.

B. Derivative of the Soft Percentile Metric

Because M_{softperc} is simply a weighted sum of all the gradient magnitudes in the image, its derivative is straightforward:

$$\frac{\partial M_{\text{softperc}}}{\partial \Delta t} = \sum_{i \in [0,S]} W_{i\text{th}} \frac{\partial G_{i\text{th}}}{\partial \Delta t} \tag{15}$$

Before proceeding to our exposure control method, we first validate our derivative formulation on a sequence recorded in an office environment. The sequence consists of images of different exposure settings of the same static scene. For each image, we calculate M_{softperc} (i.e., (10)) and then $\frac{\partial M_{\text{softperc}}}{\partial \Delta t}$ based on M_{softperc} of two consecutive images. The measured derivatives are then compared with the derivatives calculated from (15). For comparison, M_{perc} and its derivatives are also computed. For both M_{perc} and M_{softperc}, $p = 0.8$ is used.

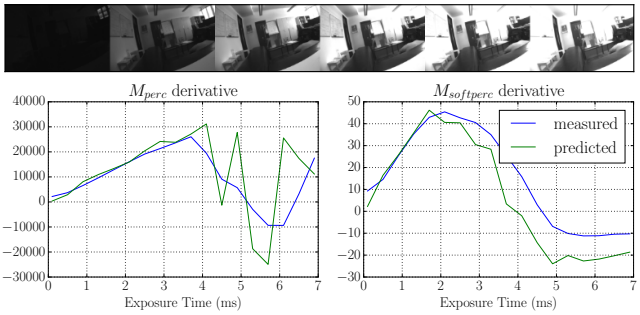


Fig. 5: Validation of the metrics derivatives. The first row shows sample images from the office sequence; the second row shows the measured and predicted derivatives of M_{perc} and M_{softperc}.

The results are shown in Fig. 5. It can be seen that the measured and predicted derivatives of M_{softperc} are close to each other. By contrast, the predicted derivatives of the M_{perc} show larger errors with respect to the measured one. This is another reason why M_{softperc} is preferred over M_{perc} in our method: the derivative of a percentile is difficult to estimate accurately. Instead of merely using the derivative from a single pixel (i.e., the pixel that has the percentile gradient magnitude), using the derivatives of all the pixels will result in a smoother and more accurate estimation.

C. Exposure Control Scheme

In Section III, we have shown that the soft percentile metric M_{softperc} is a robust indicator of the image quality. Therefore, the goal of our exposure control is to maximize M_{softperc} for future images. To achieve this goal, the exposure time is updated based on the latest image from the camera driver in a gradient ascent manner. In particular, given an image I and the corresponding exposure time Δt, the desired exposure time for the next image is calculated as:

$$\Delta t_{\text{next}} = \Delta t + \gamma \frac{\partial M_{\text{softperc}}}{\partial \Delta t}, \tag{16}$$

where the derivative of M_{softperc} is calculated by (15), and r is a design parameter to control the size of the update step. Then the new desired exposure time is sent to the camera driver and the update (16) is performed on the next image.

V. Exposure Compensation

Many VO algorithms, especially direct methods, assume that the brightness of the same part of the scene is constant over different frames. However, the change of exposure time breaks this assumption. In this section, we introduce the adaptations of two commonly used VO module—direct image alignment and direct feature matching—using both affine compensation [11], [15] and photometric compensation [1], [12], [15] and compare their performance experimentally.

A. Direct Image Alignment

Given a reference image I_r and a current image I_c, the goal of the direct image alignment is to estimate the 6 DoF motion T_{rc} (i.e., the pose of I_c in the frame of I_r). In I_r, there is a subset of pixels $S = \{\mathbf{u}_i\}$ with known depths $D = \{d_i\}$. Assuming brightness constancy, the direct image alignment estimates T_{rc} by minimizing the photometric error:

$$\mathrm{T}_{rc} = \arg\min_{\mathrm{T}} \sum_{\mathbf{u}_i \in S} [\mathrm{I}_r(\mathbf{u}_i) - \mathrm{I}_c(\mathbf{u}_i^c)]^2 \tag{17}$$

$$\mathbf{u}_i^c = \pi(\mathrm{T}^{-1}\pi^{-1}(\mathbf{u}_i, d_i)) \tag{18}$$

where $\mathbf{u}_i^c$ is the corresponding pixel of $\mathbf{u}_i$ in the current image, $\pi(\cdot)$ is the projection function that projects a 3D point into the image, and $\pi^{-1}(\mathbf{u}, d)$ backprojects a pixel $\mathbf{u}$ in the image to the corresponding 3D point, given the depth d.

When the brightness of the scene is not constant between I_r and I_c, one can use an affine transformation to model the brightness change. In this case, the direct image alignment solves the optimization problem

$$\{\mathrm{T}_{rc}, \alpha_{rc}, \beta_{rc}\} = \arg\min_{\mathrm{T},\alpha,\beta} \sum_{\mathbf{u}_i \in S} [\alpha \mathrm{I}_r(\mathbf{u}_i) + \beta - \mathrm{I}_c(\mathbf{u}_i^c)]^2. \tag{19}$$

Alternatively, if the brightness change is caused by the variation of the exposure time, we can also incorporate the photometric response function (1) into the optimization problem:

$$\mathrm{T}_{rc} = \arg\min_{\mathrm{T}} \sum_{\mathbf{u}_i \in S} [f(\frac{\Delta t_c}{\Delta t_r} f^{-1}(\mathrm{I}_r(\mathbf{u}_i))) - \mathrm{I}_c(\mathbf{u}_i^c)]^2, \tag{20}$$

where Δt_r and Δt_c are the exposure times of I_r and I_c respectively. The optimization problems (17), (19) and (20) can be solved by nonlinear least-square optimization methods such as Gauss-Newton.

B. Direct Feature Matching

Direct feature matching aims to estimate the 2D position of a feature in an image I, given an initial feature position $\mathbf{u}'$ and a reference template P of the feature. The estimation can be done by minimizing the photometric error:

$$\arg\min_{\delta\mathbf{u}} \sum_{\Delta\mathbf{u} \in P} [P(\Delta\mathbf{u}) - \mathrm{I}(\mathbf{u}' + \delta\mathbf{u} + \Delta\mathbf{u})]^2. \tag{21}$$

where $\Delta\mathbf{u}$ iterates inside the template P. The final estimation of the feature position is $\mathbf{u}' + \delta\mathbf{u}$. If an affine transformation is used to model the brightness change, (21) becomes

$$\arg\min_{\delta\mathbf{u},\alpha,\beta} \sum_{\Delta\mathbf{u} \in P} [\alpha P(\Delta\mathbf{u}) + \beta - \mathrm{I}(\mathbf{u}' + \delta\mathbf{u} + \Delta\mathbf{u})]^2. \tag{22}$$

Similar to (20), we can also use the photometric response function in the direct feature matching:

$$\arg\min_{\delta\mathbf{u}} \sum_{\Delta\mathbf{u} \in P} [f(\frac{\Delta t_c}{\Delta t_r} f^{-1}(P(\Delta\mathbf{u})) - \mathrm{I}(\mathbf{u}' + \delta\mathbf{u} + \Delta\mathbf{u})]^2 \tag{23}$$

where Δt_r is the exposure time with which the reference template is captured. Direct feature matching (21), (22) and (23) can be solved using the Lucas-Kanade algorithm [16].

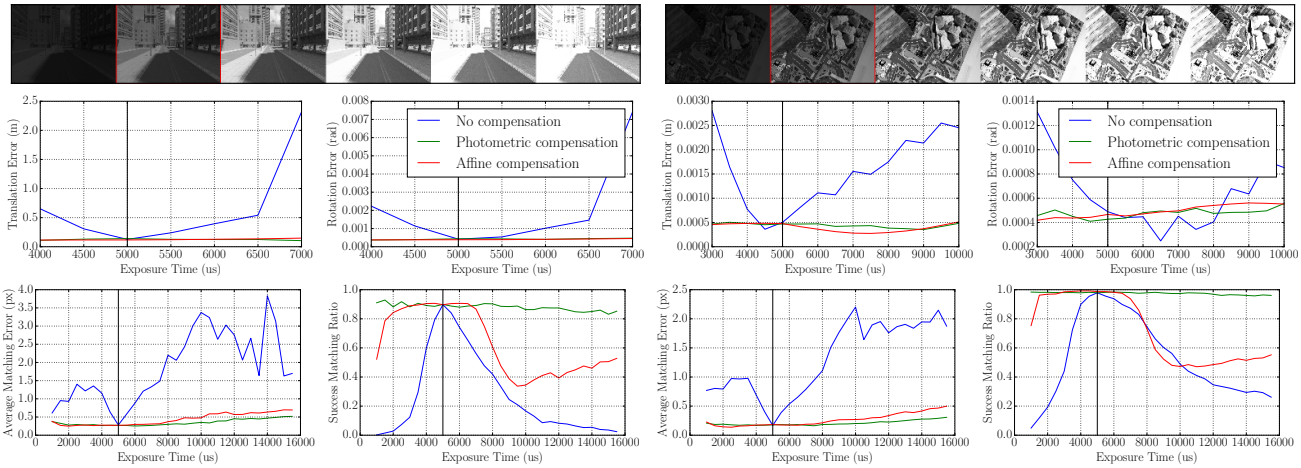


Fig. 6: Evaluation of different exposure compensation methods on synthetic datasets. Left: urban canyon dataset. Right: room dataset. The first row shows the samples of the augmented dataset, where the red square indicates the original image. The second row shows the estimation error of the direct image alignment and, the third, the success matching ratio and matching errors of the direct feature matching.

C. Evaluation

In the following, we evaluate the performance of the direct image alignment and the direct feature matching with both the exposure compensation methods (i.e., the affine compensation Eq. (19), (22) and the photometric compensation Eq. (20), (23)) on synthetic and real world datasets.

For synthetic evaluation, we use the Multi-FoV dataset [17], which contains images from two virtual scenes (urban canyon and room) with groundtruth poses and depth maps of the images. Because the dataset is rendered with a constant exposure time, we first augment the dataset using the photometric response function from a real camera (e.g., Fig. 2). In particular, we assume the exposure time of the original dataset to be a certain value, then calculate the irradiance of the scene and use the irradiance to generate images of different exposure times. Therefore, in the augmented dataset, for each *frame*, we have several *images* of different exposure times. Note that the same photometric response function is used in the photometric compensation afterwards (i.e., (20) and (23)).

To evaluate the direct image alignment, we randomly select two consecutive frames from the augmented dataset and estimate the relative transformation between the two frames. We fix one image from the first frame and use several images of different exposure times from the second frame. In our evaluation, the pixels in the small patches around the features extracted from the first image are used. The depth values of the pixels are from the ground truth depth map, and the initial pose is generated by adding a small disturbance to the ground truth pose. We measure the performance of the alignment by calculating the translation and rotation error compared to the ground truth.

The results of the direct image alignment experiment are shown in the second row of Fig. 6. It can be observed that the estimation errors of both exposure compensation methods are smaller than the situation where no compensation is applied. Interestingly, the performance of the affine compensation is similar to the photometric compensation. Note that in this experiment the response function used in the photometric compensation is perfect, in that the dataset is generated using the same function. It can be expected that on real datasets, the affine compensation will perform at least as good as the photometric compensation.

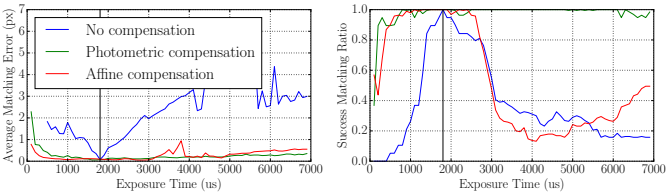


Fig. 7: Evaluation of the direct feature matching in an office environment. See Fig. 5 for image samples.

For the evaluation of the direct feature matching, we first select a random frame from the dataset and extract several FAST features from one arbitrary reference image of the frame. Then we try to match these features in all the images of the same frame. The reference templates of the features are taken from the reference image, and we add noise to the positions where the features are extracted to get the initial positions for the direct feature alignment. The success matching ratio and the final matching errors are used as performance metrics.

The results of the direct feature alignment experiment are shown in the last row of Fig. 6. Obviously, both exposure compensation methods improve the performance of the direct feature matching. Differently from the results of the direct image alignment, the photometric compensation has better performance than the affine compensation. In order to take into consideration the inaccuracy of the response function, we further evaluate the direct feature matching on the real sequence we used in Section IV-B. The results are similar, as shown in Fig. 7.

In summary, both exposure compensation methods improve the performance of the direct image alignment and the direct feature matching. Regarding the comparison between these two methods, the affine compensation performs as good as the photometric compensation in the direct image alignment, even if the latter uses a perfect photometric response function; in the direct feature matching, however, using the

photometric compensation can achieve more success matches and a better matching accuracy than the affine compensation for both synthetic and real world datasets.

VI. Experiments

In the following, we first validate our exposure method in indoor and outdoor environments. Then, we show the performance of an active VO with exposure control and compensation in real-world HDR environments.

A. Implementation Details

1) The Selection of γ: The only parameter of our algorithm is the gradient ascent rate γ. Intuitively, a large γ will make the exposure control more responsive but tend to overshoot, and a small one will have a smoother but slower behavior. By thorough outdoor and indoor experiments, we found that in general a small γ should be used for high irradiance (e.g., sunlit outdoor environment) and a large value for low irradiance (e.g., indoor environment). Therefore, we use a lookup table that maps the irradiance to γ and adjust γ at every frame. The values of the lookup table are determined experimentally.

2) Automatic Gain: In addition to the exposure time, we find it also necessary to adjust the gain of the camera. First, in extreme bright or dark scenes, it may happen that even when the camera is at its maximum/minimum exposure time, the image is still not well exposed. In such situations, the gain also needs to be adjusted. Second, the exposure time also puts a limit on the frame rate. For example, if the exposure time is too high, we can only have a low frame rate, which means that the frequency at which we can adjust the exposure time is also limited.

With the gain, denoted as g, the photometric response function (1) becomes

$$I = f(X) = f(gE\Delta t). \tag{24}$$

Obviously, to keep the image intensities constant, the exposure time should decrease/increase with same change ratio as the gain increases/descreases. In practice, we use a heuristic policy: if the exposure time is above a certain threshold, we increase the gain and decrease the exposure time accordingly, and vice versa.

3) Handling Overexposed/Underexposed Pixels: One major limitation of our method is that it exploits the gradient in the image; therefore, overexposed/underexposed pixels actually provide no information for our algorithm (i.e., the gradient and its derivative is in fact zero). If, for example, the image is totally overexposed, there is no gradient information that can be used, and then the algorithm will not adjust the exposure time at all, which is obviously not the desired behavior. We mitigate this drawback with a simple heuristic: we assign small negative derivatives (e.g., -2.0) to overexposed pixels and positive derivatives (e.g., 2.0) to underexposed ones, which forces the algorithm to react correctly to both overexposed and underexposed pixels.

B. Exposure Control

To compare our method with different camera settings, we mounted three MatrixVision Bluefox monochrome cameras in parallel on a camera rig. Each of the camera has a resolution of 752×480 pixels. The three cameras used the built-in auto-exposure algorithm, a fixed exposure time, and our exposure control algorithm, respectively. We then moved the rig in different environments and recorded the exposure time history for all the cameras.

We ran tests in 12 indoor and outdoor HDR environments (e.g., buildings under direct sunshine and shadowed areas). The fixed exposure time was hand-tuned at the start point of each test. In most of the tests, our exposure control method was able to adjust the exposure time successfully without obvious overshooting. The exposure time history in several sample sequences is shown in Fig. 8. It can observed that the exposure time variation of the built-in auto-exposure and that of our method have a similar trend. However, our method is more stable. During the test, we often observed that the exposure time of the auto-exposure could change significantly when the position of the camera changed very little (e.g., the peaks in the top-left plot of Fig. 8).

C. Active Visual Odometry

To show that combining exposure control and exposure compensation can improve the performance of VO algorithms in HDR environments, we implemented the exposure compensation methods of Section V into SVO [5]. Then we tested the adapted SVO in 10 scenes. The sequences were collected using the same three-camera setup as the previous experiment. To better show the influence of exposure control and exposure compensation separately, we tested the following configurations:

- fixed exposure time + no exposure compensation
- auto-exposure + no exposure compensation
- auto-exposure + exposure compensation
- our exposure control + exposure compensation

Based on our results on several sequences, with exposure control and compensation, the robustness and accuracy of our VO algorithm is improved. Moreover, our exposure control algorithm performs better than the auto-exposure. In the following, the results from two representative sequences are discussed in detail.

First, we show the result from a sequence in an office environment. In the sequence, the camera was first pointed toward the desk, then moved to look at the office light and lastly moved back to the initial position. Samples of the sequence are shown in Fig. 9a. To analyze the behavior of our VO algorithm in detail, we recorded the features that were tracked in each frame of the sequence. The feature tracks (frame ID vs. feature ID) are shown in Fig. 10. A dot of coordinates (x, y) in each of these plots means that feature x was tracked in frame y. A continuous vertical line indicates a feature that was persistently tracked, while a non continuous line means that the same feature was lost and then re-detected and tracked again.

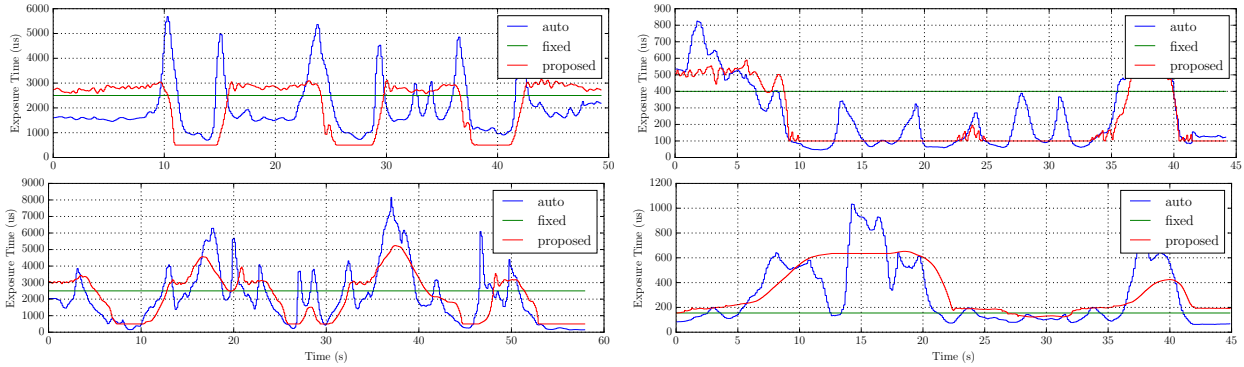
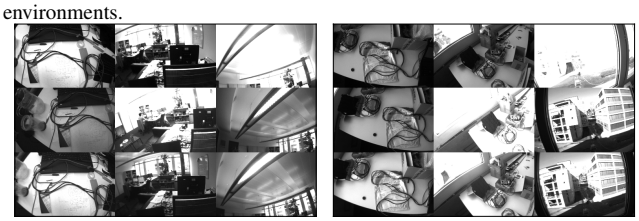


Fig. 8: Comparison of our exposure control method with the built-in auto-exposure of the camera and a fixed exposure time in both indoor and outdoor environments.



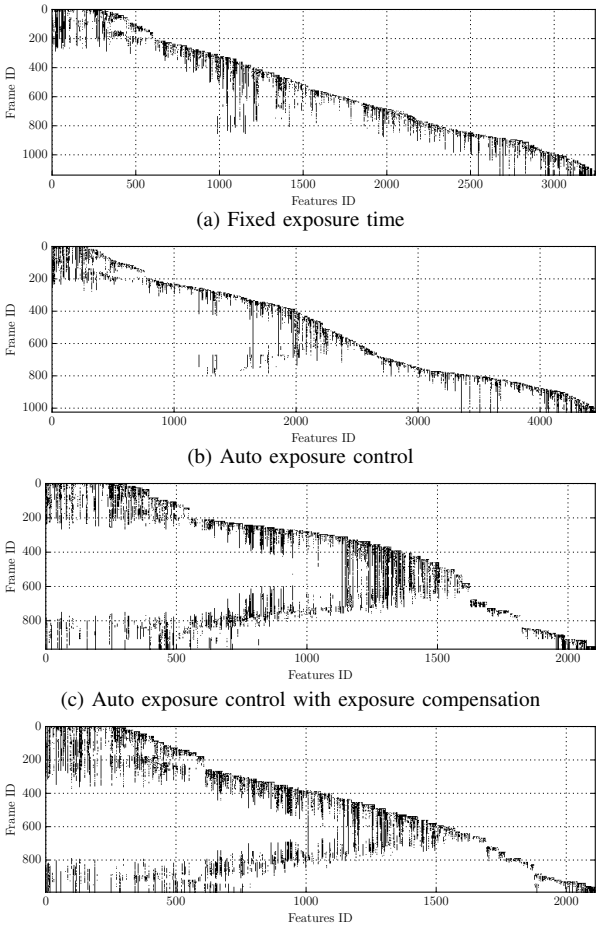
(a) Office light sequence. (b) Window sequence.

Fig. 9: Real sequences in HDR environments to test VO. First row: fixed exposure time; Second row: auto-exposure; third row: our method.

In this sequence, the adapted SVO correctly tracked the pose without loosing tracking with all the four test configurations. Comparing the configurations with exposure compensation (Fig. 10c and Fig. 10d) against the ones without (Fig. 10a and Fig. 10b), we can observe that the first two configurations present increased tracking robustness during viewpoint changes; indeed, features can be tracked longer and get more frequently re-detected. On the one hand, with a fixed exposure time, the image was badly overexposed when switching from the desk to the office light; on the other hand, when using auto-exposure without exposure compensation, the VO could not track the features well with the changing brightness.

The bad tracking in the middle also have an impact when the camera moved back to its initial position. In Fig. 10c and Fig. 10d, the VO was able to track some old features at last (i.e., the top-left area and the bottom-left area indicate the same features were tracked by both the first frames and the last frames). Obviously, this is not the case in Fig. 10a and Fig. 10b. The reason is that the aforementioned bad tracking resulted in too much drift in the last frames to correctly project the old feaures into these frames.

The result of a second test sequence is shown in Fig. 11. In this sequence, the camera was first pointed toward a desk near a window, then moved to look at the building outside the window and lastly moved to the initial position. Samples of the sequence are shown in Fig. 9b. Note that because the building was under direct sunlight at the time of recording, this sequence is more difficult than the first one. Only the configurations with exposure control were able to finish the whole sequence. Similar to our analysis of the first sequence, it can be observed that the tracking quality with our exposure control method is better than the auto-exposure.



(a) Fixed exposure time

(b) Auto exposure control

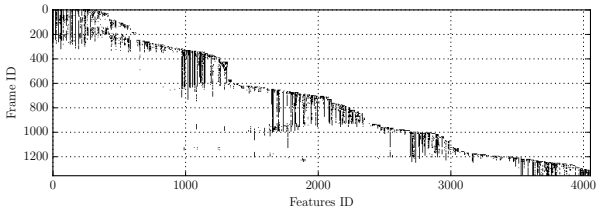
(c) Auto exposure control with exposure compensation

(d) Our exposure control with exposure compensation

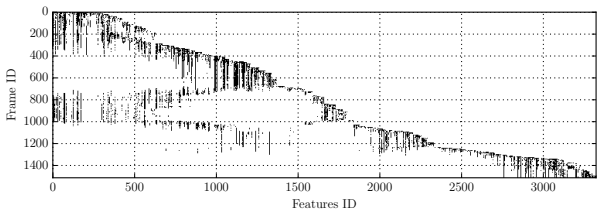
Fig. 10: Feature tracks in the office light sequence.

VII. Conclusions and Future Work

In this work, we proposed an active exposure control method to tackle this problem. We first proposed a gradient-based image quality metric and showed its robustness on various real world datasets. Then we designed a novel exposure control method, by exploiting the photometric response function of the camera, to maximize our image quality metric. We showed that our exposure control method outperforms the built-in auto-exposure of the camera in both indoor and outdoor environments. To integrate our exposure control method with VO, we introduced the adaptations for exposure compensation to a state-of-the-art algorithm. We



(a) Auto exposure control with exposure compensation



(b) Our exposure control with exposure compensation

Fig. 11: Feature tracks in the window sequence.

also experimentally compared two different exposure compensation methods and demonstrated that we can improve the robustness of VO by combining active exposure control and compensation in challenging real-world environments.

Future work would include modeling the effect of motion blur by exploiting the information from VO. Also we would like to explore the possibility to analyze the impact of the exposure time on the accuracy of VO directly.

REFERENCES

[1] S. Li, A. Handa, Y. Zhang, and A. Calway, "HDRFusion: HDR SLAM using a low-cost auto-exposure RGB-D sensor," *arXiv:1604.00895*.

[2] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *arXiv:1607.02565*, 2016.

[3] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, vol. 15, 1988, pp. 147–151.

[4] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *ECCV*. Springer, 2014, pp. 834–849.

[5] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *ICRA*, 2014, pp. 15–22.

[6] A. J. Neves, B. Cunha, A. J. Pinho, and I. Pinheiro, "Autonomous configuration of parameters in robotic digital cameras," in *Pattern Recognition and Image Analysis*. Springer, 2009, pp. 80–87.

[7] H. Kitano, M. Asada, Y. Kuniyoshi, I. Noda, and E. Osawa, "Robocup: The robot world cup initiative," in *IAA*. ACM, 1997, pp. 340–347.

[8] J. Torres and J. M. Menéndez, "Optimal camera exposure for video surveillance systems by predictive control of shutter speed, aperture, and gain," in *IS&T/SPIE Electronic Imaging*, 2015.

[9] H. Lu, H. Zhang, S. Yang, and Z. Zheng, "Camera parameters auto-adjusting technique for robust robot vision," in *ICRA*, 2010.

[10] I. Shim, J.-Y. Lee, and I. S. Kweon, "Auto-adjusting camera exposure for outdoor robotics using gradient information," in *IROS*, 2014.

[11] H. Jin, P. Favaro, and S. Soatto, "Real-time feature tracking and outlier rejection with changes in illumination," *ICCV*, vol. 1, 2001.

[12] S. J. Kim, J. M. Frahm, and M. Pollefeys, "Joint feature tracking and radiometric calibration from auto-exposure video," in *Proc. Internationl Conference on Computer Vision*, 2007.

[13] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH*. ACM, 2008, p. 31.

[14] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *ECCV*, 2006, pp. 430–443.

[15] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," 2016. [Online]. Available: http://arxiv.org/pdf/1607.02565.pdf

[16] S. Baker, R. Gross, I. Takahiro, and I. Matthews, "Lucas-kanade 20 years on: A unifying framework, Tech. Rep. CMU-RI-TR-03-01, 2003.

[17] Z. Zhang, H. Rebecq, C. Forster, and D. Scaramuzza, "Benefit of large field-of-view cameras for visual odometry," in *ICRA*, 2016.