RGB-Channel-based Illumination Robust SLAM Method

Peng Sun and Henry Y. K. Lau
Department of Industrial and Manufacturing Systems Engineering, the University of Hong Kong, Hong Kong
Email: sunpeng@connect.hku.hk, hkylau@hku.hk

***Abstract*—This paper provides an illumination robust direct monocular simultaneous localization and mapping (SLAM) method, which takes advantage of the RGB channel to enhance the lighting change invariance of the input frame series. By linearly combining the RGB channel, some illumination-insensitive components in the colour space are extracted and represented by three indicators in this paper. An optimization model is then provided to minimize the errors in these indicators using a Kalman filter (KF). These indicators serve to update the frames by keeping only the illumination-insensitive components. The illumination robust visual SLAM method based on these enhanced frames is thereby offered. In addition, the gradient magnitude is utilized to improve the distinctiveness of each pixel in the frame. Experiments on both artificial and natural datasets show that the provided method has a better illumination robustness than the state-of-the-art direct SLAM method.**

***Index Terms*—visual SLAM, illumination robustness, colour constancy**

I. Introduction

In the past few years, the visual SLAM method has attracted the attention of many researchers and has been used in many applications such as self-driven car, quadcopter control and 3D reconstruction. The state-of-the-art visual SLAM/odometry methods such as the LSD-SLAM [1], ORB_SLAM [2] and DSO[3] methods achieve high precision in both pose tracking and scene reconstruction. However, performance degradation occurs, especially in the direct SLAM method, when the lighting changes rapidly. The reason is that the frame-to-frame matching accuracy is significantly affected by the illumination conditions. Thereby, the challenging problem of improving the illumination robustness is raised, and the related experiments are necessary and crucial in terms of extending the application range of SLAM methods.

This paper provides an approach to reduce changes in the input frame series caused by changes in lighting intensity and to, thereby, improve the illumination robustness of the SLAM method. To perform the reduction, we extract the illumination-insensitive components from every frame, which retain similar values under changing lighting conditions. The frames are subsequently adjusted by replacing the original intensities with these components to resist lighting changes. Running on the enhanced frame series, the proposed SLAM method has an improved illumination robustness.

We make use of the RGB colour channels to extract the illumination-insensitive components. Studies of colour channel analysis show that some illumination irrelevant information hides in the different intensities of different colour channels. First, we follow and extend Maddern et al.'s work[4], which shows that the intensity of each colour channel is represented by a polynomial of illumination-relevant terms and illumination-irrelevant terms, to extract the illumination-insensitive components by linearly combining the red, green and blue channels. Furthermore, we perform a statistical study on the Amsterdam Library of Object Images (ALOI)[5] and find that the darkest one among the RGB channels has a robustness to light source temperature changes. Keeping the darkest channel is an effective approach to reduce the error caused by light temperature. Based on the above knowledge, three indicators that have superior performance under different illumination conditions are generated in this paper.

For the purpose of merging these three indicators together to get a better estimate of the illumination-insensitive components, we assign a weight to each indicator and utilize an optimization model to calculate the weights based on the illumination conditions. A colour-balance-based method is used to partly estimate the weights, and a KF method is then employed to dynamically adjust the weights during the robot movement. After the optimization is performed, the gradient magnitude is then applied on every pixel to improve their distinctiveness.

Several experiments are performed on the New Tsukuba Stereo Dataset (NTSD)[6] and the UR-10-based natural scenes dataset to compare our method to the LSD-SLAM method[1] in terms of the tracking accuracy and 3D-point-reconstruction performance. The results show that the provided method has a better illumination robustness.

The main contributions and innovations of this paper are as follows.

1. Taking advantages of the RGB colour channels, multiple illumination-insensitive indicators are provided. In contrast to the single indicator method, our method provides a better illumination robustness.

Manuscript received July 27, 2019; revised November 1, 2019.

©2019 Journal of Automation and Control Engineering

doi: 10.18178/joace.7.2.61-69

2. Utilizing an optimization model of merging the above indicators, better estimates of the illumination-insensitive components are calculated and dynamically adjusted with lighting changes.
3. The use of gradient magnitude improves the distinctiveness of every pixel.

The remainder of this paper is organized as follows. Section II discusses related work. The derivation and definition of the three indicators are considered in III-A. The optimization model is described in III-B. The gradient magnitude is shown in III-C. Section IV presents the experimental results, and section V is the conclusion.

II. Related Work

A. Visual SLAM

The visual SLAM method utilizes only a camera to gain the depth information of objects and then completes the localization and mapping task. Based on the different image matching methods, the visual SLAM methods are classified into two groups: the direct method, which finds corresponding points by directly comparing intensity values, and the feature-based method, which extracts the feature points first and then finds corresponding points among these feature points. In general, the feature-based methods, such as the ORB-SLAM [2] and ORB-SLAM2 [7] methods, have a better illumination robustness than direct methods because the feature extraction intrinsically filters out some illumination-sensitive components. However, direct methods such as the LSD-SLAM [1] method provide more detailed reconstruction models by taking into account the whole image rather than only a small selection of feature points. Some studies have improved the illumination robustness of the direct SLAM methods. The NID-SLAM [8] method utilizes the global intensity distribution to reduce local intensity changes and, thereby, obtain illumination robustness. However, the iterative histogram computation results in a low operation rate. The DSO [3] method provides a photometric calibration method to resist lighting changes, but this calibration inevitably needs extra work. Therefore, we focus on the imaging method that directly obtains illumination-irrelevant intensities to improve the illumination robustness.

B. Illumination Robust Imaging

Some methods balance accuracy and computational complexity by taking advantage of the scene structure. Son et al.'s work [9] proposes a high-speed method that utilizes the vanishing point and the region of interest to estimate the possible lighting changes in the scene. Simo-Serra et al.'s work [10] takes advantage of the heat fusion map to embed the image as a physical shape, which thereby resists illumination changes by taking into account position information. Kim et al.'s work [11] provides an image matching method that utilizes the Mahalanobis distance cross-correlation, which is illumination invariant, to measure the similarity between two images. These methods perform well, but their application range is limited because they require specific input data.

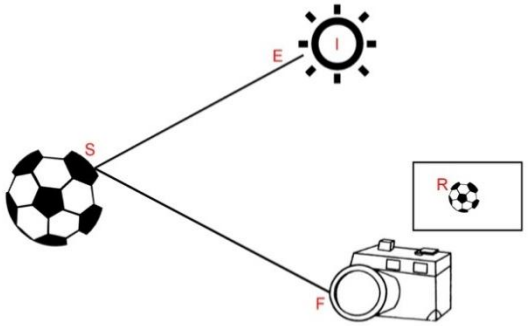

Figure 1. A demonstration of the imaging process.

Some other methods utilize extra devices to achieve illumination robustness. Hu et al.'s work [12] reduces the intensity changes caused by lighting by using Lidar data as the reference. Maddern et al.'s work [13] provides a method to resist day/night lighting changes by computing the prior 3D structure and then inferring the illumination invariant area based on structural knowledge. These methods have a high performance under challenging illumination conditions, but they require expensive equipment to obtain the extra data.

A new path to solve this problem is to build an imaging model and exclude the illumination-relevant components. Ratnasingam and Collins [14] discuss an imaging model using different wavelengths of light to calculate the illumination-insensitive components. Then, Maddern[4] fixes the wavelength to specific values so that an RGB camera could be used to obtain the data.

III. Direct Monocular SLAM With Illumination-insensitive Components

In this section, we describe the details of our illumination robust SLAM method, especially the derivation and definition of the illumination-insensitive indicators, the optimization model and the gradient magnitude approach. We adopt the framework of [1] to describe the frames and the map.

A. Illumination-insensitive Indicators

The direct SLAM method usually estimates the relative pose by minimizing the photometric error. Given an image $\boldsymbol{I}: \boldsymbol{\Omega} \to \mathbb{R}^+$ and the depth map $\boldsymbol{D}: \boldsymbol{\Omega} \to \mathbb{R}^+$, every pixel $\boldsymbol{p_i} \in \boldsymbol{\Omega}$ is transformed, with the associated depth information $\boldsymbol{D_r}(\boldsymbol{p_i})$, to its corresponding point q_i on the new frame by the camera pose $\boldsymbol{\xi}$. Here, we adopt the transformation function in [1] to perform this operation.

$$\boldsymbol{q_i} = \boldsymbol{\omega}(\boldsymbol{p_i}, \boldsymbol{D_r}(\boldsymbol{p_i}), \boldsymbol{\xi}) \tag{1}$$

The camera pose $\boldsymbol{\xi}$ is then iteratively estimated by minimizing the photometric error between the reference frame $\boldsymbol{I_r}$ and the current frame $\boldsymbol{I_r}$ using the following equation:

$$\boldsymbol{\underset{\xi}{arg\,min}} \sum_{\boldsymbol{p_i} \in \boldsymbol{\Omega}} (\boldsymbol{I_r}(\boldsymbol{p_i}) - \boldsymbol{I_c}(\boldsymbol{q_i}))^2 \tag{2}$$

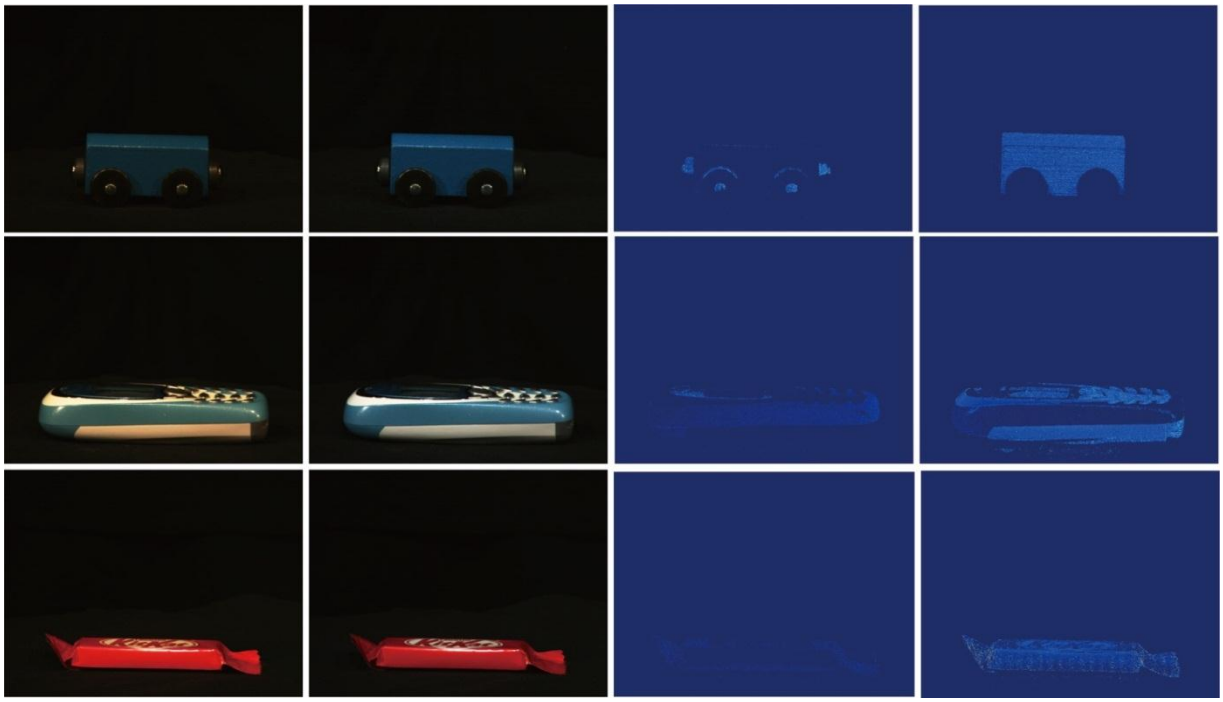

Figure 2. A demonstration of the illumination robustness of the darkest channel. Columns 1 and 2 show two images under different lighting conditions. Columns 3 and 4 show the intensity change values of the darkest channel and the brightest channel, respectively. Brighter parts in the image imply a more intense transformation in intensity. We can see the darkest channel changes less when the illumination changes

In our method, the illumination-insensitive value (IIV) is utilized to measure the photometric error.

$$\underset{\xi}{\boldsymbol{arg\,min}} \sum_{\boldsymbol{p_i}\in\boldsymbol{\Omega}} (\boldsymbol{IIV_r}(\boldsymbol{p_i}) - \boldsymbol{IIV_c}(\boldsymbol{q_i}))^2 \tag{3}$$

where $\boldsymbol{IIV_r}(\boldsymbol{p_i})$ is a synthesis of three illumination-insensitive indicators of the reference frame and $\boldsymbol{IIV_c}(\boldsymbol{p_i})$ for the current frame. The derivation of indicators is discussed in the following part.

The illumination-insensitive indicators are derived from the imaging model. Fig. 1 shows a general imaging process in which a light source emits a beam of light onto the object, and the light is reflected by the surface of the object and is finally perceived by a camera. Ratnasingam and Collins[14] provided a model to represent this process:

$$\boldsymbol{R^{x,E}} = \boldsymbol{a^x n^x I} \int \boldsymbol{S^x}(\boldsymbol{\lambda})\boldsymbol{E}(\boldsymbol{\lambda})\boldsymbol{F}(\boldsymbol{\lambda})\boldsymbol{d\lambda} \tag{4}$$

where $\boldsymbol{x}$ represents an arbitrary point on the image, which is also associated with a point in the 3D real world; $\boldsymbol{R^{x,E}}$ is the intensity of the point under the specific spectral distribution $\boldsymbol{E}$ of the light source; $\boldsymbol{a^x}$ is the input direction of the light; $\boldsymbol{n^x}$ is the normal of the surface near the corresponding real world point of $\boldsymbol{x}$; $\boldsymbol{\lambda}$ is the wavelength of different components in the light; $\boldsymbol{I}$ and $\boldsymbol{E}(\boldsymbol{\lambda})$ represent the intensity and spectral density per wavelength of the light source, respectively; $\boldsymbol{S^x}(\boldsymbol{\lambda})$ is the surface reflectively at point $\boldsymbol{x}$; and $\boldsymbol{F}(\boldsymbol{\lambda})$ represents the sensitivity of the camera to different wavelengths.

The model of equation (4) is precise but hard to use because it contains too many parameters. With the help of some other studies, a more practicable equation is derived from (4). First, Jiang et al.'s work[15] shows that $\boldsymbol{F}(\boldsymbol{\lambda})$ has a narrow response range so we can use a Dirac delta function to represent it. Second, according to Finlayson's et al.'s work[16], $\boldsymbol{E}(\boldsymbol{\lambda})$ obeys Planck's law so it can be rewritten in a simple form. Finally, [14] and [4] provide a derivation of (4).

$$\boldsymbol{log\, R_i} = \boldsymbol{log}(\boldsymbol{G^x I}) + \boldsymbol{log}\left(\boldsymbol{hc^2 \lambda_i^{-5} S_i}\right) - \frac{\boldsymbol{hc}}{\boldsymbol{k_B T \lambda_i}} \tag{5}$$

where $\boldsymbol{h}$ is the Planck constant, $\mathbf{c}$ is the speed of light, $\boldsymbol{k_B}$ is the Boltzman constant, $\boldsymbol{G^x} = \boldsymbol{a^x n^x}$, and $\boldsymbol{T}$ is the correlated colour temperature of the light source. There are three illumination-relevant values: the $\boldsymbol{G^x}$ of the light direction, the $\boldsymbol{I}$ of the light intensity and the $\boldsymbol{T}$. Thereby, we can obtain the illumination-insensitive indicators by removing the above items, which is done by taking into account the RGB colour channel.

We adopt [4] to obtain the first illumination-insensitive indicator in this paper.

$$\boldsymbol{\mathcal{I}_1} = \mathbf{log}\left(\boldsymbol{R_g}\right) - \boldsymbol{\alpha}\,\mathbf{log}(\boldsymbol{R_r}) - (\mathbf{1}-\boldsymbol{\alpha})\,\mathbf{log}(\boldsymbol{R_b}) \tag{6}$$

where α is estimated by the following equation to remove the illumination-relevant terms:

$$\frac{\boldsymbol{hc}}{\boldsymbol{k_B T \lambda_g}} - \frac{\boldsymbol{\alpha hc}}{\boldsymbol{k_B T \lambda_r}} - \frac{(\mathbf{1}-\boldsymbol{\alpha})\boldsymbol{hc}}{\boldsymbol{k_B T \lambda_b}} = \mathbf{0} \tag{7}$$

which can be simplified to

$$\frac{\mathbf{1}}{\boldsymbol{\lambda_g}} - \frac{\boldsymbol{\alpha}}{\boldsymbol{\lambda_r}} - \frac{\mathbf{1}-\boldsymbol{\alpha}}{\boldsymbol{\lambda_b}} = \mathbf{0} \tag{8}$$

where λ_g, λ_r and λ_b are the camera's intrinsic parameters representing the most sensitive wavelengths of the camera's green, red, and blue channels, respectively. These parameters can be measured by a monochromator and a spectrophotometer[17]. Therefore, $\mathcal{I}_1$ contains only the illumination-irrelevant terms and can be calculated using only the measurable parameters.

However, using this indicator alone has some limitations. First, as discussed above, the measurement of λ_i requires extra equipment, which can be used only in the lab environment. Second, λ_i has its own error, which cannot be avoided by using only one indicator. Third, with only one single indicator, the different reflectance values could be confused.[18] Therefore, we derive two other λ_i -irrelevant indicators to overcome the above limits. One of them is defined as

$$\boldsymbol{\mathcal{I}_2} = \frac{\boldsymbol{log(R_r) - log(R_g)}}{\boldsymbol{log(R_b) - log(R_g)}}$$

$$= \frac{\boldsymbol{log\left(\frac{S_r}{S_g}\right) + log[E(\lambda_r)] - log[E(\lambda_g)]}}{\boldsymbol{log\left(\frac{S_b}{S_g}\right) + log[E(\lambda_b)] - log[E(\lambda_g)]}} \tag{9}$$

$$= \frac{\boldsymbol{log[\frac{S_r}{S_g}] - log[\frac{E(\lambda_r)}{E(\lambda_g)}]}}{\boldsymbol{log[\frac{S_b}{S_g}] - log[\frac{E(\lambda_b)}{E(\lambda_g)}]}}$$

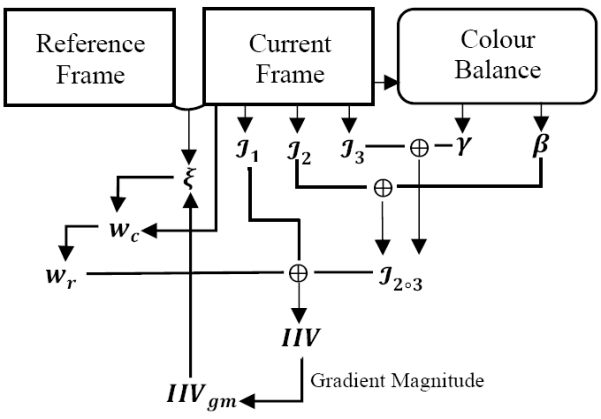

Figure 3. The demonstration of the optimization process.

Assuming that $\boldsymbol{E(\lambda)}$ is constant, $\boldsymbol{\mathcal{I}_2}$ contains only the illumination-irrelevant components. Actually, in short-term travel, $\boldsymbol{E(\lambda)}$ usually keeps the same value because the type of light source rarely changes. For example, in a daylight scene, the light source is always the sun, and only the direction of light affects the illumination. Therefore, $\mathcal{I}_2$ is useful to resist frame-to-frame illumination changes.

If $\boldsymbol{E(\lambda)}$ changes, we provide another indicator to improve the illumination robustness.

$$\boldsymbol{\mathcal{I}_3 = min\, R_i, i = r, g, b} \tag{10}$$

The definition of this indicator means that we choose the darkest channel of the pixel as $\mathcal{I}_3$. To find the $\boldsymbol{E(\lambda)}$-insensitive indicator, we take advantage of the dataset of ALOI[5], which contains 1000 sets of image taken under 12 different light temperatures. A measurement of the average intensity changes of separated channels under different light temperatures is performed. The result, as in Fig. 2, shows that the darkest channel tends to keep the same value during the light change. The average intensity change of the darkest channel is 23 compared to 52 for the brightest channel. Occasionally, the opposite case occurs, but we could still use this indicator in most of situations because a smaller intensity value usually implies a smaller S_i so that it changes less under the imaging model expressed in equation (1).

With the above three indicators, we can define $\boldsymbol{IIV}$.

$$\mathbf{IIV}(\mathbf{p_i}) = \sqrt{(\mathbf{w}\boldsymbol{\mathcal{I}_1})^2 + [(\mathbf{1}-\mathbf{w})\boldsymbol{\mathcal{I}_{2\circ3}}]^2} \tag{11}$$

where $\boldsymbol{w}$ is a weight to measure the reliability of the two indicators $\boldsymbol{\mathcal{I}_1}$ and $\boldsymbol{\mathcal{I}_{2\circ3}}$, and $\boldsymbol{\mathcal{I}_{2\circ3}}$ is a linear combination of $\boldsymbol{\mathcal{I}_2}$ and $\boldsymbol{\mathcal{I}_3}$. In practice, $\boldsymbol{\mathcal{I}_1}$ and $\boldsymbol{\mathcal{I}_{2\circ3}}$ have different contributions to the $\boldsymbol{IIV}$ because they result in different types of error and describe different aspects of the illumination-insensitive components. Thus, $\boldsymbol{w}$ is applied to them to represent the difference. $\boldsymbol{\mathcal{I}_{2+3}}$ is defined as

$$\boldsymbol{\mathcal{I}_{2\circ3}} = \frac{\boldsymbol{\beta\mathcal{I}_2 + \gamma\mathcal{I}_3}}{\boldsymbol{\beta + \gamma}} \tag{12}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the weights based on the change of the colour balance. Since $\boldsymbol{\mathcal{I}_2}$ is based on the assumption of a constant $\boldsymbol{E(\lambda)}$ and $\boldsymbol{\mathcal{I}_3}$ compensates for the error caused by $\boldsymbol{E(\lambda)}$ changes, $\mathcal{I}_{2\circ3}$ provides a better estimate of illumination-insensitive components. The calculations of $\boldsymbol{w}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are discussed in III-B.

In conclusion, in this section we demonstrated how $\boldsymbol{IIV}$ works in the pose tracking process and provided three illumination-insensitive indicators to help build $\boldsymbol{IIV}$. We also gave the weighted relations between $\boldsymbol{IIV}$ and the three indicators.

B. Kalman-filter-based Optimization Model

In this section, we demonstrate the estimation of the weights, which help to merge the three indicators. First, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are calculated by functions of colour balance. Then, $\boldsymbol{w}$ is calculated with a KF-based method. A demonstration of the entire optimization process is shown in Fig. 3.

$\boldsymbol{\mathcal{I}_2}$ is provided under the assumption of a constant $\boldsymbol{E(\lambda)}$, which is normal in our practice. However, in the case of an abrupt change of $\boldsymbol{E(\lambda)}$, the accuracy of $\boldsymbol{\mathcal{I}_2}$ will be

rapidly reduced. In contrast, $\mathcal{I}_3$ resists $E(\lambda)$ changes but causes confusion due to decreased contrast. Therefore, we utilize the two weights β and γ to measure the different accuracies of $\mathcal{I}_2$ and $\mathcal{I}_3$ under variant light temperatures. First, we define a colour balance change between two frames as

$$\Delta c = -\left(e^{\frac{\sum_{p\in\Omega} I_r^R(p)}{\sum_{p\in\Omega} I_r^B(p)}} - e^{\frac{\sum_{q\in\Omega} I_c^R(q)}{\sum_{q\in\Omega} I_c^B(q)}}\right) \tag{13}$$

where $I_r^R(p)$ is the red channel intensity of point p in the reference frame r, R indicates the red channel, and $I_r^B(p)$, $I_c^R(q)$ and $I_c^R(q)$ are similar. We assume that the distance (both position and orientation) between two frames is short enough so that only a small amount of pixels change. In this situation, the ratio of red channel to blue channel can be approximately estimated using the global sum value. Here, Δc has two aspects. First, it is literally the ratio of red channel to blue channel. Second, it is the light temperature change, which is affected mainly by the red and blue components. The dual character of Δc makes it useful in estimates of both β and γ.

Because $\mathcal{I}_2$ obeys the assumption of an invariant spectral distribution, β is defined as a Boolean function:

$$\beta = \begin{cases} 1, & |\Delta c| \le t \\ 0, & |\Delta c| > t \end{cases} \tag{14}$$

where t gives a tolerance of the error in the computation of Δc. We estimate it by calculating the average error of Δc in the New Tsukuba Stereo Dataset[6].

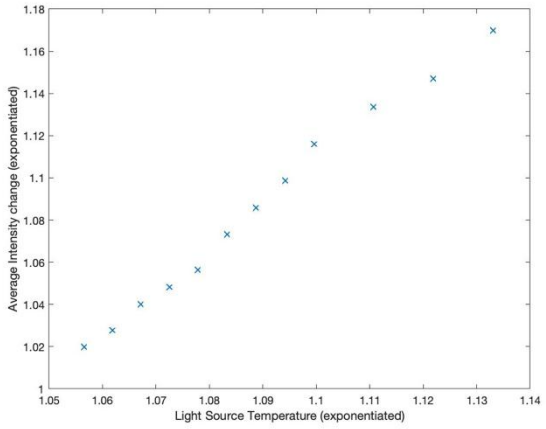

Figure 4. The linear relation between the intensity deviation of the darkest channel and the light source temperature. Note that both the axes are exponentiated.

The estimation of γ is based on an analysis of ALOI [5]. Fig. 4 shows the linear relation between the intensity deviation of the darkest channel and the light source temperature change. Note that $I_r(p_i)$ is normalized to resist the light intensity change. Thus,

$$e^{\mathcal{I}_3 - I_r^{dark}(p)} = k\,\Delta c + b \tag{15}$$

where $I_r^{dark}(p)$ is considered to be the true intensity of p. We define γ as

$$\gamma = \frac{\mathcal{I}_3 - log(k\,\Delta c + b)}{\mathcal{I}_3} \tag{16}$$

where k and b are calculated using the least squares method to fit the point pairs in ALOI with the variant light temperatures.

Thus far, with the estimates of β and γ, the combined indicator $\mathcal{I}_{2\circ 3}$ is generated with robustness to both light intensity deviations and spectral distribution changes.

$\mathcal{I}_1$ and $\mathcal{I}_{2\circ 3}$ describe the illumination-insensitive components in different views, so that the magnitude of vector $(\mathcal{I}_1, \mathcal{I}_{2\circ 3})$ is illumination robust and, moreover, more distinguishable than any single indicator of these two. However, their respective reliabilities vary according to different scenes. For example, $\mathcal{I}_1$ follows the assumption of a narrow response range so that its accuracy is degraded when the camera is sensitive to a wide band of wavelengths. $\mathcal{I}_3$ remains more constant under the high contrast of different channels, which means $\mathcal{I}_{2\circ 3}$ works better in more "colourful" scenes. Thus, we assign a weight to adjust the magnitude according to the scene and provide a KF-based method to dynamically estimate the weights during travel.

w is associated only with key frames because the scene change is a long-term process and successive adjustment is not necessary. An iteration method is applied to update w frame by frame. For every new frame, we first utilize the reference weight w_r to calculate IIV and the camera pose ξ. The following equation is then solved to find a better estimate of w_i at specific point p:

$$\underset{p_i\in\Omega_D}{IIV(p_i)} = \sqrt{w_i^2\mathcal{I}_1(q_i)^2 + (1-w_i)^2\mathcal{I}_{2\circ 3}(q_i)^2} \tag{17}$$

where q_i is calculated by equation (2), and Ω_D signifies the points with depth information $D(p)$. Then, the measurement of the entire current frame w_c is calculated as the average of all w_i, and the variant of w_c is recorded as

$$\sigma_c = \sqrt{\frac{1}{\Omega_D - 1}\sum_{p_i\in\Omega_D}(w_i - w_c)^2} \tag{18}$$

We use the Kalman filter to merge w_c into w_r and treat σ_c as a 1-D covariance of the observation noise.

Figure 5. Examples of New Tsukuba Stereo Dataset and our UR-10-based dataset. The left group of images is the New Tsukuba Stereo Dataset. Clockwise from the top left: Daylight, flashlight, lamp and fluorescent light. The right group of images is our UR-10-based dataset, and each image is extracted from four independent tracks.

TABLE I. SUCCESSFUL TRACKING RATE DURING DIFFERENT ILLUMINATION

Track 1 / Track 2	Daylight		Flashlight		Lamp		Fluorescent	
	LSD-SLAM	Our	LSD-SLAM	Our	LSD-SLAM	Our	LSD-SLAM	Our
Daylight	**96.1**	95.2	45.2	**52.1**	54.3	**60.7**	72.4	**87.8**
Flashlight	37.2	**46.5**	92.2	**94.6**	0.0	0.0	**49.1**	47.4
Lamp	24.6	**40.0**	32.7	**37.1**	89.6	**93.5**	78.6	**89.7**
Fluorescent	71.2	**86.7**	47.3	**52.4**	55.6	**67.2**	**96.5**	96.1

C. Gradient Magnitude Method

The use of $\boldsymbol{IIV}$ inevitably causes a degradation of distinctiveness, which means multiple pixels with different intensities may share a similar $\boldsymbol{IIV}$ value. To compensate for this degradation, we employ the gradient magnitude method, which shows great illumination robustness and an acceptable computational complexity according to [19]. After updating the frame with $\boldsymbol{IIV}$, we further calculate the gradient magnitude of pixel p in a 3×3 neighbourhood. The final illumination-insensitive value of a single pixel is defined as follows:

$$\boldsymbol{IIV}_{\boldsymbol{gm}} = \sqrt{\frac{\boldsymbol{\partial IIV(p)}^2}{\boldsymbol{\partial x}} + \frac{\boldsymbol{\partial IIV(p)}^2}{\boldsymbol{\partial y}}} \tag{19}$$

IV. EVALUATION

We evaluate the illumination robustness of our method by a comparison with the LSD-SLAM method[1]. The evaluation is performed on three databases: the New Tsukuba Stereo Dataset[6] with synthetic scenes, the natural indoor scenes under flashlight illumination collected by a UR 10 robot, which has the ground truth camera position (GT data), and another natural indoor scene under variant illumination (No-GT data).

The New Tsukuba Stereo Database provides four tracks of the same path under different synthetic illumination conditions. To provide a natural dataset, we utilize the UR 10 to collect four tracks under flashlight illumination. Some examples of the two datasets can be seen in Fig. 5. We also provide a dataset with no ground truth data, which includes 100 sets of images under 7 different illumination conditions to extend the range of illumination change types. Some samples can be seen in Fig. 6.

The experiments evaluate three aspects of the provided method. First, we perform our method and the method of Maddern et al.[4] on the No-GT data to prove that the multiple indicators method is superior to the single indicator method. We apply the two methods on images under different illumination conditions and then perform the SIFT[20] method on the enhanced images to match the feature points. A larger volume of correctly matched points shows the better illumination robustness of the method.

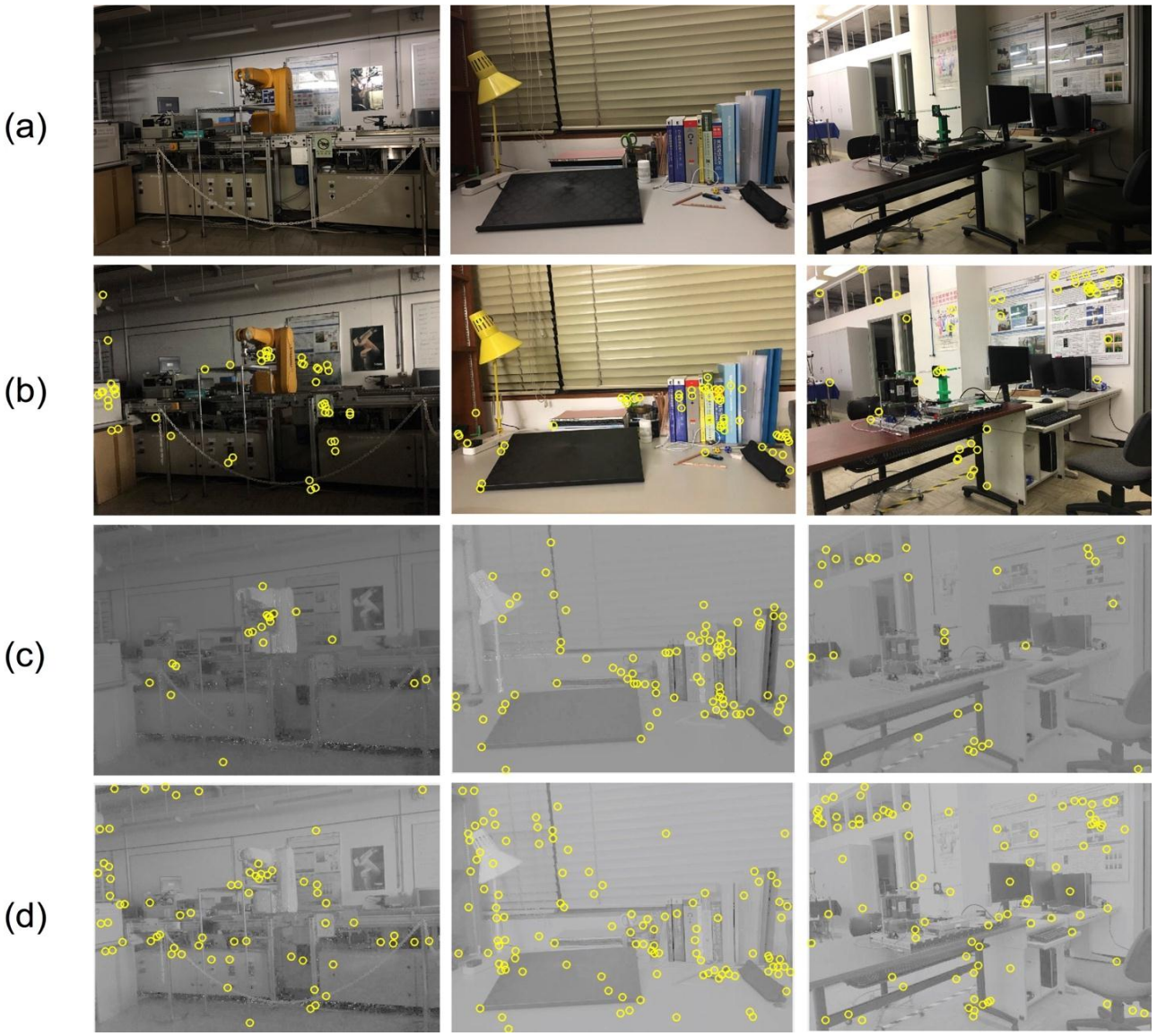

Figure 6. Qualitative results of feature point extraction with different methods. (a) Image under a specific illumination condition. (b) Image taken at the same view under different illumination conditions, with the SIFT feature points. (c) Feature points extracted by SIFT with Maddern's work. (d) Feature points extracted by SIFT with our work. The yellow circles imply the feature points that are kept during the RANSAC process.

Second, we test the localization robustness of our method using the New Tsukuba Stereo Dataset[6]. In the test, we perform the target method on two tracks under different illumination conditions and record the successful retrieval rate of the frames in the second track against the key-frames in the first track, which is similar to the approach in Pascoe et al.[8].

In the past few years, the SLAM method has focused on not only the tracking accuracy but also the mapping quality. Therefore, we design the experiment to evaluate the mapping performance of our method in a challenging environment. We compare our method with the LSD-SLAM method on the flashlight scene of the New Tsukuba Stereo Dataset and the GT data and record the number of reconstructed points.

A. Multiple Indicators Performance

Three methods are compared in this experiment: the SIFT method, the SIFT method with a single indicator[4] and the SIFT method with our multiple indicators. For every set in the GT data, we first extract the feature points on the two images under different illumination conditions and then match these points to record the number of correctly matched point pairs. A RANSAC method is employed to filter out incorrect matching. Fig. 6 shows some qualitative results. We record the average matched number of every set of images, which is shown in Fig. 7.

The results show that although the single indicator method improves the SIFT feature quality under varying illumination conditions, the multiple indicator method has a better illumination robustness. The reason is that the single indicator reduces the distinctiveness of the pixel, and the multiple indicators help to recover it in a way.

©2019 Journal of Automation and Control Engineering

TABLE II. NUMBER OF RECONSTRUCTED POINTS AND TRACKING ACCURACY

Method \ Track	NTSD		GT data track 1		GT data track 2		GT data track 3		GT data track 4	
	Point Number	Accuracy (mm)	Point Number	Accuracy (mm)	Point Number	Accuracy (mm)	Point Number	Accuracy (mm)	Point Number	Accuracy (mm)
LSD-SLAM	8245	99.2	3221	84.3	6795	63.5	3542	82.3	3196	87.2
Ours	**9459**	**87.1**	**4231**	**81.1**	**9742**	**45.2**	**4672**	**77.1**	**4593**	**84.1**

B. Localization Performance

We apply the LSD-SLAM method and our method on the New Tsukuba Stereo Dataset to measure the retrieval rate. Table I shows the percentage of frames located in the second track against the key frames in the first track. Our method is superior to the LSD-SLAM method in most of the experiments, which shows its high illumination robustness.

C. Mapping Performance

We utilize both the synthetic and natural methods to evaluate the mapping quality of our method. The flashlight scene is a typical challenging environment because the movement of the light source causes uneven local changes of intensity. We apply the LSD-SLAM method and our method on five tracks, one of which is extracted from the New Tsukuba Stereo Dataset and four of which are from the GT data. The number of reconstructed points is shown in Table II along with the tracking accuracy in millimetres. We can see that our method reconstructed more points than the LSD-SLAM method, while retaining a high localization accuracy.

D. Limitations

Despite the high illumination robustness, the provided method still has some limitations. First, the computational complexity of our method is higher than that of the LSD-SLAM method due to the extra computation of $\boldsymbol{IIV}$. Even when we try not to involve the weight optimization in the iteration calculation of $\boldsymbol{\xi}$, this method can only perform at 10 Hz on a desktop CPU. Second, the provided method cannot handle extreme lighting changes such as complete darkness, and failure occurs when the flashlight illuminates only a small area.

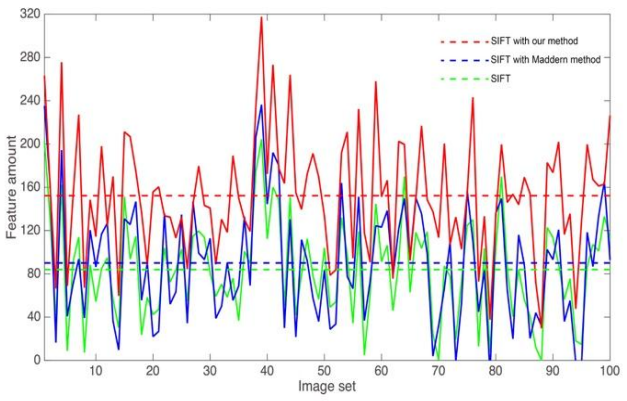

Figure 7. The number of reconstructed points. The X-axis is the index number of the image set. The Y-axis is the average number of reconstructed points in a specific set. The horizontal dashed line indicates the average level among all the sets.

V. CONCLUSION

This paper provides an illumination robust monocular direct SLAM method that takes advantage of RGB channel information to find the illumination-insensitive components and, thus, achieve pixel-level invariance under different lighting conditions. We propose three indicators that consist of illumination-invariant terms and provide an optimization method to merge them into the same framework. The main contributions of this work are as follows. (1) In contrast to the single indicator method, our method utilizes multiple indicators to reduce the confusion of contrast associated with decreases in colour information. (2) The KF-based optimization method makes it possible to dynamically adjust the weight of the indicators so that the illumination robustness is improved under different lighting conditions. (3) The gradient magnitude improves the distinctiveness of the pixels and, thus, increases the successful matching rate between the points. The provided method shows an advantage in illumination robustness compared to that of the state-of-the-art direct SLAM method. We believe our method can extend the application range of the SLAM method in terms of rapidly changing illumination conditions.

REFERENCES

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

[2] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147-1163, 2015.

[3] J. Engel, V. Koltun, and D. Cremers, "Direct sparse odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611-625,2018.

[4] W. Maddern, et al. "Illumination invariant imaging: Applications in robust vision-based localisation, mapping and classification for autonomous vehicles," in *Proc. of the Visual Place Recognition in Changing Environments Workshop, IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China. 2014.

[5] J. M. Geusebroek, G. J. Burghouts, and A. W. Smeulders, "The Amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103-112,2005.
[6] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005.
[7] 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.
[8] G. Pascoe, et al. "NID-SLAM: Robust monocular SLAM using normalised information distance," in *Proc. Conference on Computer Vision and Pattern Recognition*, 2017.
[9] J. Son, et al., "Real-time illumination invariant lane detection for lane departure warning system," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1816-1824,2015.
[10] E. Simo-Serra, C. Torras, and F. Moreno-Noguer, "DaLI: deformation and light invariant descriptor," *International Journal of Computer Vision*, vol. 115, no. 2, pp. 136-154, 2015.
[11] S. Kim, et al., "Mahalanobis distance cross-correlation for illumination-invariant stereo matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 11, pp. 1844-1859, 2014.
[12] X. Hu, F. S. A. Rodriguez, and A. Gepperth, "A multi-modal system for road detection and segmentation," in *Proc. 2014 IEEE Intelligent Vehicles Symposium Proceedings*, 2014.
[13] W. Maddern, A. D. Stewart, and P. Newman. "LAPS-II: 6-DoF day and night visual localisation with prior 3D structure for autonomous road vehicles," in *Proc.* Intelligent Vehicles Symposium Proceedings, 2014.
[14] S. Ratnasingam and S. Collins, "Study of the photodetector characteristics of a camera for color constancy in natural scenes," *JOSA A*, vol. 27, no. 2, pp. 286-294, 2010.
[15] J. Jiang, et al. "What is the space of spectral sensitivity functions for digital color cameras?" in *Proc. 2013 IEEE Workshop on Applications of Computer Vision (WACV)*, 2013.
[16] G. D. Finlayson, et al., "On the removal of shadows from images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59-68, 2006.
[17] J. Nakamura, "Image sensors and signal processing for digital still cameras," 2016: CRC press.
[18] M. Ebner, "Color constancy," vol. 6. 2007, John Wiley & Sons.
[19] S. Park, T. Schöps, and M. Pollefeys, "Illumination change robustness in direct visual slam," in *Proc. 2017 IEEE international conference on robotics and automation (ICRA)*. 2017.
[20] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

Peng Sun received a B.A. degree in Electronic Business from Zhongnan University of Economics and Law and a master's degree in Computer Science from Shandong University. His research interest includes computer vision and SLAM.

He is currently a PhD candidate in the Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong.

Henry Y.K. Lau received a B.A. degree in engineering science and a D.Phil. degree in robotics from the University of Oxford, Oxford, U.K. His research interests include artificial intelligence, in particular artificial immune systems, intelligent automation for material handling, and virtual and augmented reality systems.

He was a Croucher Foundation Research Fellow with the University of Oxford Robotics Research Group and a Visiting Lecturer with the Brasenose College at Oxford, where he was involved in teaching engineering science. He worked in industry for many years as a Systems Engineer and a Section Manager with the U.K. Atomic Energy Authority and AEA Technology plc., where he was involved in projects with a focus on bespoke telerobotics systems and advanced automation systems for the nuclear industry in decommissioning and waste management. He is currently an Associate Professor with the Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong.