Intrinsic Image Based Illumination Robust Tracking Method

SUN Peng
The University of Hong Kong
The University of Hong Kong, Hong Kong
sunpeng@connect.hku.hk

Henry Y.K. Lau
The University of Hong Kong
The University of Hong Kong, Hong Kong
hyklau@hku.hk

Abstract

This paper presents an illumination robust SLAM tracking method which is underpinned by an efficient intrinsic image generation approach. We transform the color image to an intrinsic image to mitigate the matching error caused by the lighting change. In contrast to the state-of-the-art work, we offer an optical flow based optimization procedure to efficiently estimate a more precise transformation parameter in the untrained image series. We evaluate the method in the scene under various illumination conditions and show that it is more robust than using the original color image.

1. Introduction

In the past decade, the topic of Simultaneously Localization and Mapping (SLAM) attracted a lot of attention therefore a vast number of methods have been provided. Despite the increased precision and reduced computational complexity, the state-of-the-art method often relies on the assumption of constant illumination, which unfortunately is usually not achievable in practice. To solve the contradiction, the research in the field of color constancy endeavors to estimate the invariance to the lighting changes thus benefit the tracking task in terms of reducing the photometric error.

In this paper, we follow [1] to transform the original color image to an intrinsic image, which remains invariant to diverse illumination because of it depends only on the intrinsic parameter of the camera and the reflectivity of the object in the scene. The transformation parameter plays a crucial role in generating the intrinsic image however it is a challenge to estimate it. The state-of-the-art work estimates this parameter with either a trained image series which is inconvenient [2] or an inaccurate means [1, 3]. Instead of manually pairing corresponding points in the contiguous frames, we simultaneously perform the parameter estimation and the corresponding points pairing to make it available with untrained frames. An optimization is executed by minimizing the difference of intrinsic value between corresponding points to offer a higher precision.

We also take advantage of the optical flow to make available an expedited optimization with the Levenberg-Marquardt method.

The novelties of this method are:

1) In contrast to the global similarity based SLAM method like [4, 5], the provided method requires no extra iterative computation for illumination invariance.
2) In contrast to the current state-of-the-art intrinsic image generation method, the provided method needs no trained image series to estimate the transformation parameter.
3) The employment of the optical flow expedites the optimization of the transformation parameter and increases the precision.

2. Illumination robust SLAM tracking

2.1. Intrinsic image generation

We follow [1] to calculate the color composition of a pixel with the following equation:

$$\rho_i = \log(c_i) - \log(\sqrt[3]{c_r c_g c_b}) \qquad (1)$$

Where c_i is the intensity of the color channel of Red, Green or Blue. The composition vector is then projected to a plane:

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{2}{3}} & -\sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{6}} \\ 0 & -\sqrt{\frac{1}{2}} & \sqrt{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \rho_r \\ \rho_g \\ \rho_b \end{bmatrix} \qquad (2)$$

[1] proves that the same color under different illumination satisfies a linear relation in this plane, an illustration is shown in Fig. 1. So that an illumination invariant value is estimated by:

$$g = E_1 cos\theta + E_2 sin\theta \qquad (3)$$

θ is the transformation parameter. The intrinsic image is generated by replacing the intensity of every pixel with g.

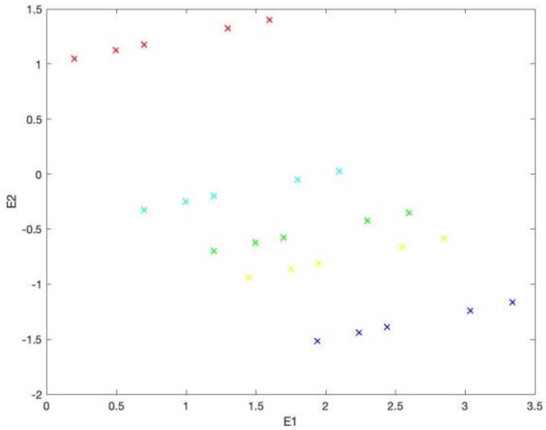

Figure 1. The illustration of different colors in the E_1-E_2plane. The color of the points indicates its real color, the different position is caused by illumination change. The illustration shows that the coordinates of the same color under different illumination falls along parallel lines so that the direction of the line implies an invariance to illumination change.

2.2. Estimation of transformation parameter

We estimate the transformation parameter by minimizing the difference of intrinsic value between two corresponding points in the contiguous frames with the following equation:

$$\arg\min_{\theta} \underbrace{\|g(x,y) - g(x+dx, y+dy)\|^2}_{r^2} \tag{4}$$

Where (x,y) and $(x+dx, y+dy)$ are two corresponding points. We use the Levenberg-Marquardt method to solve this minimization problem so that the partial differential of $\frac{\partial r}{\partial \theta}$ is necessary. For this reason, we take the advantage of the Lucas-Kanade optical flow (LK method) to offer the relation between (dx, dy) and θ, based on which the $\frac{\partial r}{\partial \theta}$ can be calculated.

Noticing that the intrinsic value remains invariance during the camera movement, we get the following equation:

$$g(x+dx, y+dy, t+dt)_{|\theta^*} = g(x,y,t)_{|\theta^*} \tag{5}$$

From the Taylor expansion of the left term in eq. 5, the following equation is derived:

$$\frac{\partial g}{\partial x}\frac{dx}{dt} + \frac{\partial g}{\partial y}\frac{dy}{dt} = -\frac{\partial g}{\partial t} \tag{6}$$

By substituting ep. 3 into ep. 6 and rewriting it in the matrix form, we get the following equation of the camera movement speed:

$$[cos\theta \quad sin\theta] P \begin{bmatrix} u \\ v \end{bmatrix} = -[cos\theta \quad sin\theta] T \tag{7}$$

Where

$$P = \begin{bmatrix} \frac{\partial E_1}{\partial x} & \frac{\partial E_1}{\partial y} \\ \frac{\partial E_2}{\partial x} & \frac{\partial E_2}{\partial y} \end{bmatrix}; T = \begin{bmatrix} \frac{\partial E_1}{\partial t} \\ \frac{\partial E_2}{\partial t} \end{bmatrix}; u = \frac{dx}{dt}; v = \frac{dy}{dt} \tag{8}$$

Assuming that k points in the neighbor area of (x,y) have the same motion, we get the following equation:

$$RA \begin{bmatrix} u \\ v \end{bmatrix} = -Rb \tag{9}$$

Where

$$A = \begin{bmatrix} P_1 \\ \vdots \\ P_k \end{bmatrix}; b = \begin{bmatrix} T_1 \\ \vdots \\ T_k \end{bmatrix}; R = \begin{bmatrix} [cos\theta \quad sin\theta]^T{}_1 \\ \vdots \\ [cos\theta \quad sin\theta]^T{}_k \end{bmatrix}^T \tag{10}$$

The camera movement speed $(u,v)^T$ is subsequently calculated by the least square method:

$$\begin{bmatrix} u \\ v \end{bmatrix} = -(A^{\mathrm{T}} R^T R A)^{-1} A^T R^T R b \tag{11}$$

Thus far we get the relation between (dx, dy) and θ. Therefore the $\frac{\partial r}{\partial \theta}$ is computable. We minimize the eq.4 to obtain the transformation parameter θ and subsequently transform the color image to intrinsic image. The tracking method is performed with the intrinsic images to improve the illumination robustness.

3. Experiment

We evaluate our method on the Oxford RobotCar Dataset [6]. Four tracks on the same route under different illumination conditions are selected for the comparison. We compare our method with Finlayson's work[1] and Corke's work [2] to measure the precision of the generated intrinsic image. We use the SIFT method to pair the corresponding points then calculate the similarity of every point pair in the intrinsic image generated with the methods under comparison. The result is shown in Table I, it indicates that our method has a high accuracy than the other two methods.

Table I: The comparison of intrinsic generating methods

Methods	Track 1	Track 2	Track 3	Track 4
Ours	**92.12%**	**88.73%**	**89.49%**	**94.52%**
Finlayson's	66.43%	62.20%	59.43%	64.68%
Corke's	85.72%	81.33%	83.14%	86.92%

4. Conclusion

In this paper, we provide an illumination robust SLAM tracking method. This method takes advantage of the intrinsic image to achieve an improved performance under various lighting conditions and the reduced computational complexity. In contrast to the state-of-the-art intrinsic image generation method, our method does not require trained image to estimate the transformation parameter.

Furthermore, owing to the employment of the optical flow, the provided method offers a more accurate intrinsic image.

References

[1] Finlayson, G.D., M.S. Drew, and C. Lu, "Entropy minimization for shadow removal," International Journal of Computer Vision, 85(1): p. 35-57,2009.

[2] Corke, P., et al. "Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation." in 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2013. IEEE.

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

[4] Pascoe, G., et al. "NID-SLAM: Robust Monocular SLAM Using Normalised Information Distance." in Conference on Computer Vision and Pattern Recognition. 2017.

[5] Wang, X., M. Christie, and E. Marchand. "Multiple layers of contrasted images for robust feature-based visual tracking." in 2018 25th IEEE International Conference on Image Processing (ICIP). 2018. IEEE.

[6] Maddern, W., et al., "1 year, 1000 km: The Oxford RobotCar dataset," The International Journal of Robotics Research, 36(1): p. 3-15,2017.