International Journal of Control, Automation and Systems 17(X) (2019) 1-9
http://dx.doi.org/10.1007/s12555-018-0199-2
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Patch-based Stereo Direct Visual Odometry Robust to Illumination Changes 基于patch的光照稳定性

Jae Hyung Jung, Sejong Heo, and Chan Gook Park*

Abstract: In this paper, we present a patch-based direct visual odometry (DVO) that is robust to illumination changes at a sequence of stereo images. Illumination change violates the photo-consistency assumption and degrades the performance of DVO, thus, it should be carefully handled during minimizing the photometric error. Our approach divides an incoming image into several buckets, and patches inside each bucket own its unique affine illumination parameter to account for local illumination changes for which the global affine model fails to account, then it aligns small patches placed at temporal images. We do not distribute affine parameters to each patch since this yields huge computational load. Furthermore, we propose a prior weight as a function of the previous pose in a constant velocity model which implies that the faster a camera moves, the more likely it maintains the constant velocity model. Lastly, we verify that the proposed algorithm outperforms the global affine illumination model at the publicly available micro aerial vehicle and the planetary rover dataset which exhibit irregular and partial illumination changes due to the automatic exposure of the camera and the strong outdoor sunlight, respectively.

Keywords: Affine illumination model, direct visual odometry, micro aerial vehicle, nonlinear optimization, rover navigation.

1. INTRODUCTION

Estimating an ego-motion of a camera has been one of the most challenging tasks for a camera mounted moving platform in global navigation satellite system (GNSS) denied environment. One way to tackle this issue is to use visual odometry (VO) which was coined its name owing to its similarity to wheel odometry (WO). VO estimates a relative 6-DOF pose between consecutive images and incrementally obtains its pose and does not suffer from error accumulation caused by wheel slips, a tremendous disadvantage in WO [1]. VO is widely used in a robot navigation because of cost and space effectiveness of cameras. VO system was successfully implemented in NASA's Mars Exploration Rover (MER). MER's VO system tracked corner features at Martian terrain and estimated relative poses between an incoming pair of images by the stereo camera [2]. VO in a micro aerial vehicle (MAV) application can be found in [3] which exploited VO with a downward-looking camera attached to the MAV.

VO can be divided into two types so-called indirect VO (IVO) and direct VO (DVO), depending on which information is provided to the cost function in the optimization problem. IVO [4] minimizes a reprojection error defined as a difference between a feature measurement and an estimated feature location, while DVO [5] estimates camera's pose by minimizing a photometric error which is intensity difference among consecutive images. DVO is known to outperform IVO in motion blurred and featureless condition since it does not utilize feature information but pixel intensities directly in images [5]. However, DVO has a substantial weakness that it is vulnerable to illumination changes in a sequence of images. This is because DVO assumes that every object in the world has the same intensity regardless of viewer's position that is known as the Lambertian surface. The assumption is invalid under practical conditions where sudden and irregular illumination changes are prevalent in sequences of images attributable to automatic exposure and gain of a camera or albedo change that is caused by an irregular reflection under outdoor sunlight.

The above illumination issues are common in image related problems which directly exploit pixel intensities such as DVO and target tracking algorithms. Specifically, [6] employed a photometric normalization method for the face tracking algorithm under illumination changes. Also, [7] proposed the algorithm that recognizes the current

Manuscript received March 28, 2018; revised June 19, 2018; accepted August 16, 2018. Recommended by Associate Editor Kang-Hyun Jo under the direction of Editor Euntai Kim. This work was supported by the Ministry of Science and ICT of the Republic of Korea through the Space Core Technology Development Program under Project NRF-2018M1A3A3A02065722.

Jae Hyung Jung, Sejong Heo, and Chan Gook Park are with the Department of Mechanical and Aerospace Engineering/Automation and System Research Institute, Seoul National University, Daehak-dong, Gwanak-gu, Seoul 08826, Korea (e-mails: {lastflowers, jjong80, chan park}@snu.ac.kr).
* Corresponding author.

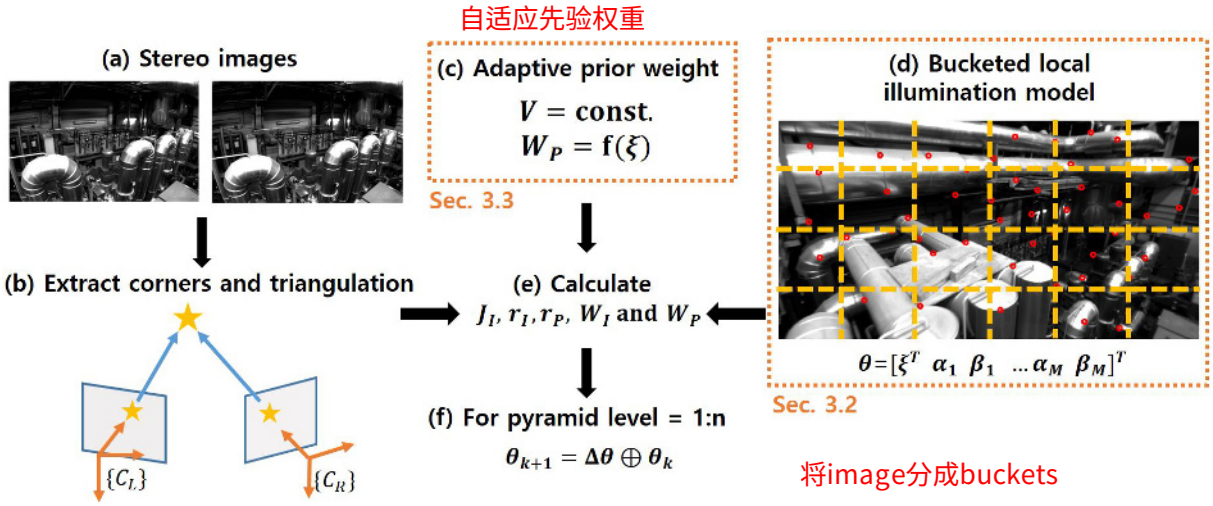


Fig. 1. The overview of the proposed algorithm: (a, b) reconstruct features based on the static stereo baseline, ${}^{C_L}T_{C_R}$ (e) calculate the Jacobian matrix, J_I, residuals, r_I, r_P and weighting matrices, W_I, W_P according to (c) the constant velocity model with the adaptive prior weight and (d) the bucketed local illumination model under (f) the image pyramid loops.

illumination level of the environment, and selects one of the pre-built maps with different brightness for the localization. In VO, [22] formulated the local illumination parameters for each planar patch and then marginalized out by projecting to the null space in the EKF framework. Also, [8, 9] estimated global illumination parameters with the camera's poses employing the illumination affine model [10] and the single illumination offset, respectively from RGB-D images. In [11], the global affine illumination parameter was estimated in the alternative fashion that fixes the pose and the illumination parameter in turn to deal with the outliers. To take account of local illumination changes, [12] selected planar patches sharing the same affine illumination parameters [10] and jointly optimized the photometric error for the pose and the affine parameter in RGB-D cameras. However, the images should have enough planar patches to obtain reliable motion estimation in [12] which limits application domain into an indoor environment where artificial structures make rich planar patches for proper motion estimation.

The main contribution of this paper is threefold. First, we propose a patch-based DVO which is robust to illumination changes at stereo camera images employing the *bucketed local illumination model* (Fig. 1(d)). In our model, patches centered at feature points have the same affine illumination parameters within the buckets in the image as shown in Fig. 1(d) where patches are marked as the red dots. Therefore, the proposed model requires less computational cost than the local illumination model which augments its state vector per a patch. Also, the generated patches enable the proposed method to work in a more general environment, since it does not require any artificial planar patches, while accounting for not only global light changes but also local light changes. Second, we propose the *adaptive prior weight* (Fig. 1(c)) as a function of the previously converged motion in a constant velocity motion model framework. This reflects a physical intuition that the faster a camera moves, the harder it is to change a velocity—we assign a weight to the constant velocity model according to the previous motion. In cases where a motion is huge, the proposed method improves estimation accuracy, as will be seen in Section 4. Lastly, we show experimental evidence that the proposed algorithm outperforms global illumination model in the lunar-like terrain dataset [14] with strong outdoor sunlight and the MAV dataset [15] where camera's automatic exposure and gain make sudden and partial illumination changes throughout image sequences.

The remainder of this paper is organized as follows: in Section 2, we give simple mathematical preliminaries about DVO. Next, Section 3 describes the proposed local illumination change model with its mathematical formulation, and the adaptive temporal weight in the sparse image alignment problem. In Section 4, the proposed algorithm is evaluated in real-world datasets of MAV and rover navigation where sudden and irregular illumination changes are prevalent. Finally, Section 5 summarizes the conclusion of this paper.

2. MATHEMATICAL PRELIMINARIES

2.1. Notation and relative pose 符号和相对pose

DVO estimates relative poses between a current camera frame, $\{C_2\}$ and a previous camera frame, $\{C_1\}$ and concatenates them to obtain global pose referenced at a global 连锁

frame, $\{G\}$. The relative pose, $\xi \in se(3)$ is defined as

$$\xi = \begin{bmatrix} {}^{C_2}v_{C_1}^T & {}^{C_2}w_{C_1}^T \end{bmatrix}^T \Delta t, \tag{1}$$

where Δt is timestamp interval between $\{C_1\}$ and $\{C_2\}$, and v and w are linear and angular velocity, respectively. Throughout this paper, the left superscript refers to a referenced frame and the right subscript refers to an object frame. ξ is mapped to Special Euclidean group, $SE(3)$ through an exponential mapping,

$${}^{C_2}T_{C_1} = \exp(\hat{\xi}) \in SE(3), \tag{2}$$

where T is a rigid body transformation matrix and the hat operator, $\hat{\cdot}$ is defined as follows with the skew-symmetric matrix operator, $[\cdot_\times]$ 斜对称矩阵

$$\hat{\xi} = \begin{bmatrix} [{}^{C_2}w_{C_1\times}] & {}^{C_2}v_{C_1} \\ 0 & 1 \end{bmatrix}. \tag{3}$$

2.2. Camera projection model

In this paper, the standard pinhole camera model is adopted and the projection model for a j-th feature viewed at $\{C_1\}$, is defined as

$$\begin{bmatrix} u_{f_j} \\ v_{f_j} \end{bmatrix} = \Pi({}^{C_1}P_{f_j}) = \begin{bmatrix} \frac{f_u {}^{C_1}X_{f_j}}{{}^{C_1}Z_{f_j}} + c_u \\ \frac{f_v {}^{C_1}Y_{f_j}}{{}^{C_1}Z_{f_j}} + c_v \end{bmatrix}, \tag{4}$$

where ${}^{C_1}P_{f_j} = \begin{bmatrix} {}^{C_1}X_{f_j} & {}^{C_1}Y_{f_j} & {}^{C_1}Z_{f_j} \end{bmatrix}^T$ is the location of the j-th feature, Π is the projection model, $f_{u,v}$ and $c_{u,v}$ are a focal length and a principle point, respectively.

A warping is a transformation of a pixel location from one image plane to another according to a relative motion between $\{C_1\}$ and $\{C_2\}$, and the warping function, $w(\cdot)$ for the i-th pixel, x_i is defined as

$$w(\xi, x_i) = \Pi(g({}^{C_2}T_{C_1}(\xi),\ \Pi^{-1}(x_i))), \tag{5}$$

where $g(T,\ P) = P' \in \mathbb{R}^3$ is a rigid body motion mapping. In other words, the warping is a projection of the same feature to different image planes of camera frame according to the their relative pose.

2.3. Photometric error

DVO assumes that every object in a image has Lambertian surface property, i.e., photo-consistency assumption, therefore, the photometric error for the i-th pixel is defined as

$$r_i(\xi) = I_1(x_i) - I_2(w(\xi, x_i)) \tag{6}$$

where $I_1 : \mathbb{R}^2 \to \mathbb{R}$ and $I_2 : \mathbb{R}^2 \to \mathbb{R}$ are previous and current grayscale images, respectively [5]. DVO minimizes the squared sum of photometric errors with respect to the relative pose.

$$\xi^* = \underset{\xi}{\operatorname{argmin}} \sum_i^n r_i^2(\xi). \tag{7}$$

Equation (7) is also known as the image alignment problem in the sense that the 6-DOF pose, ξ aligns a pair of temporal images.

3. PROPOSED DVO ALGORITHM

An overview of the proposed algorithm is presented in Fig. 1. First of all, features (corner, blob, etc.) are extracted from a pair of stereo images in the bucketed manner like in [13], and a matching algorithm finds a correspondence for each feature, then, the matched features are reconstructed by two-view structure-from-motion optimization. Small patches are generated, which are centered at the extracted features. Next, the prior pose yields the prior residual and the adaptive prior weight. We augment the state vector with the affine illumination parameters and jointly estimate the relative pose and the illumination parameters based on the *bucketed local illumination model*. The current estimate of the augmented state vector iteratively computes the photometric error term. To obtain a good initial guess for the estimator, we employ the coarse-to-fine scheme as in [5]. After solving the optimization problem, the obtained relative pose is concatenated to calculate the global pose of the camera, ${}^GT_{C_k}$. The detailed explanation for the algorithm is given in the following subsections.

3.1. Problem formulation

The main objective of this paper is to solve the photometric minimization problem, (7) to obtain the relative pose between two temporally successive images. However, (7) might converge to a false minimum or even diverge under a severe brightness change environment or a large motion of a camera. In other words, if the brightness change affects the temporal images or the overlapping region between the consecutive images is not large enough, the nonlinearity of the cost function is increased so that the estimator might fail. To account for the issues, we solve the following modified optimization problem,

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \begin{bmatrix} r_I^T & r_P^T \end{bmatrix} \begin{bmatrix} W_I & 0 \\ 0 & W_P \end{bmatrix} \begin{bmatrix} r_I \\ r_P \end{bmatrix} \tag{8}$$

that is equivalent to the maximum a posterior (MAP) estimator of $p(\xi|r, \xi_P)$ where $p(r) = p(r_{1:n})$ with the independent and identically distributed (iid) assumption for the measurements and zero mean of photometric errors [17]. We denote a prior as subscript P in the following sections, for instance, ξ_P stands for a pose prior.

In (8), θ is the state vector composed of the relative pose and affine illumination parameters defined as follows:

$$\theta = \begin{bmatrix} \xi^T & \alpha_1 & \beta_1 & \cdots & \alpha_M & \beta_M \end{bmatrix}^T \in \mathbb{R}^{6+2M}. \tag{9}$$

The affine illumination parameter, α_l and β_l are contrast and brightness changes, respectively [10], and M stands

for the total number of buckets. In (8), W_I is an image weighting matrix which is determined by the distribution of the photometric error. For example, [17] proposed several weighting matrices such as the T-distribution weighting matrix. Also, r_I is vectorized illumination compensated photometric error, and r_P is the residual from the prior pose,

$$r_I = \begin{bmatrix} r_{1,\text{affine}} & r_{2,\text{affine}} & \cdots & r_{n,\text{affine}} \end{bmatrix}^T, \tag{10}$$

$$r_P = \xi_P - \xi. \tag{11}$$

A choice of interesting pixels, i.e., the elements of r_I in (10) is a crucial strategy in terms of computational efficiency and estimation accuracy. For instance, [5] uses all pixels whose depth are valid for a motion tracking, and [16] reduces interesting image region to pixels with non-negligible intensity gradient. Also, [3] proposes the sparse image alignment that aligns patches centered at a sparse set of features. We adopt the sparse image alignment proposed by [3] because we do not triangulate all pixels but corner features from the static stereo pair. In addition to this, the constant depth assumption in a small patch is reasonable while reducing computational burden.

3.2. Bucketed local illumination model

The number of the illumination parameters in (9) is directly related to the dimension of the state vector. Therefore, assigning the parameters to each pixel is computationally impractical. For instance, in a 640×480 resolution image, its state vector has 614,406 dimensions. On the other hand, a global illumination model as in [8] where $M = 1$ assumes that whole pixels in an image undergo the same intensity changes with the single pair of the parameter. However, this assumption is violated in practical applications due to partial illumination changes on the image. To address this issue, we propose a local illumination model that accounts for both global and local brightness changes. 为了处理违反照片一致性假设的突发和局部照明变化，我们将唯一的仿射照明参数分配给每个桶中的稀疏斑块。

To deal with sudden and partial illumination changes that violate the photo-consistency assumption, we distribute the unique affine illumination parameters to the sparse patches in each bucket. Note that a bucket in an image is a region divided by the grids as shown in Fig. 1(d), for instance, $M = 24$ in case of Fig. 1(d). Accordingly, similar to [8], the photometric error is modified as follow

$$r_{i,\text{affine}}(\theta) = I_1(x_i) - [(\alpha_l + 1) I_2(w(\xi, x_i)) + \beta_l], \tag{12}$$

and the linearized photometric error of (8) is

$$r_I(\theta_{k+1}) \cong r_I(\theta_k) + J_I(\theta_k)\Delta\theta, \tag{13}$$

where the augmented state vector yields the following modified Jacobian matrix,

$$J_I = -\begin{bmatrix} \frac{\partial r_{I,\text{affine}}}{\partial \xi} & \frac{\partial r_{I,\text{affine}}}{\partial \alpha_1} & \cdots & \frac{\partial r_{I,\text{affine}}}{\partial \alpha_M} & \frac{\partial r_{I,\text{affine}}}{\partial \beta_M} \end{bmatrix}, \tag{14}$$

$$\frac{\partial r_{I,\text{affine}}}{\partial \xi} = \begin{bmatrix} (\alpha_1 + 1) & \cdots & (\alpha_M + 1) \end{bmatrix}^T J_\xi(\xi_k), \tag{15}$$

where J_ξ is the Jacobian matrix of the photometric error that is computed by the chain rule from (5) as follows:

$$J_\xi(\xi_k) = \left.\frac{\partial I_2}{\partial \Pi}\right|_{\Pi_k} \left.\frac{\partial \Pi}{\partial g}\right|_{g_k} \left.\frac{\partial g}{\partial {}^{C_2}T_{C_1}}\right|_{T_k} \left.\frac{\partial {}^{C_2}T_{C_1}}{\partial \xi}\right|_{\xi_k}, \tag{16}$$

where $\partial I_2/\partial\Pi$ is an image gradient, $\partial\Pi/\partial g$ is a derivative of a pixel position to its 3-D position, $\partial g/\partial T$ is a derivative of a 3-D position of a feature to a rigid body motion, and $\partial T/\partial\xi$ is a derivative of a rigid body motion to a twist. Then, a normal equation is obtained after solving the first order necessary condition for (8),

$$\Delta\theta = (J_I^T W_I J_I + W_P)^{-1}(-J_I W_I r_I + W_P r_P). \tag{17}$$

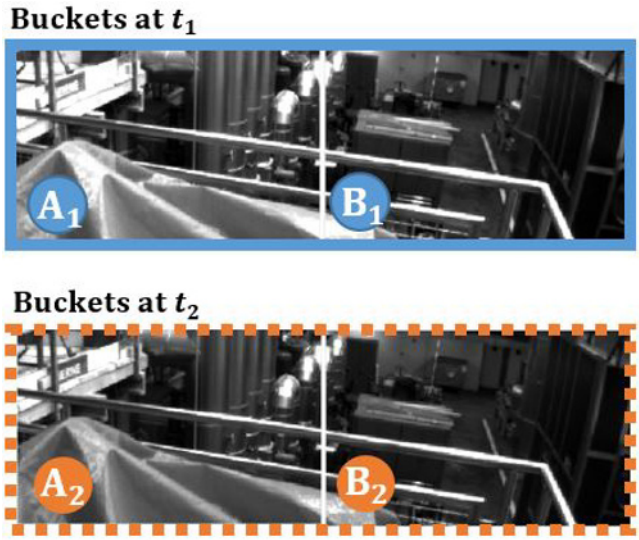
Lastly, the relative pose is updated through exponential and logarithm mapping with the hat operator defined in (2),

$$\hat{\xi}_{k+1} = \log(\exp(\Delta\hat{\xi}) \cdot \exp(\hat{\xi}_k)). \tag{18}$$

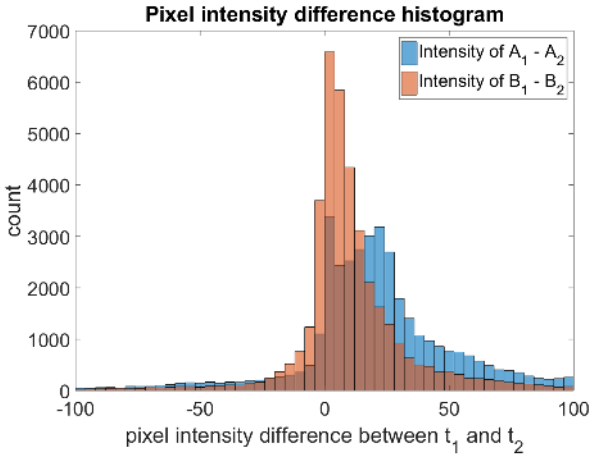
We suppose that each patch located in the same bucket possesses its own affine parameter to account for local illumination changes, and name this model as the *bucketed local illumination model*. Note that since a pair of the parameter adds two additional states to the state vector, θ in (9), we do not distribute affine parameters to each patch but to patches that belong to each bucket for reducing computational burden while accounting for local illumination changes. Also, pixels in each patch share the same parameters because of the fact that the small patches, e.g., 3×3, can be approximated locally tangent plane in a smooth surface showing similar intensity changes to illumination changes [10]. Also, in a temporal sequence of images, we suppose that patches stay within the same buckets without loss of generality because camera's frame rate (10-60 fps) is high enough to make patches stay in their bucket in general. By assigning a large number of small patches rather than few large patches as in [12], we do not need planar patch fitting and all depth value inside each patch. Also, since the proposed algorithm does not align artificial planar patches, it can operate in both outdoor and indoor environments.

3.3. Adaptive prior weight

The prior weight, W_P indicates how certain we believe the constant velocity model in the total cost function, (8), i.e., the inverse of prior's uncertainty in the MAP estimator. To deal with this weighting matrix in the absence of additional sensor like IMU or odometry, [18] conducts parametric studies and obtains the best constant diagonal weighting matrix at its given dataset in a heuristic manner. However, in this paper, we suppose the system model



(a) Sample bucket pairs.



(b) Intensity difference histogram.

b1和b2几乎没区别，但是a1和a2有较大的差值

Fig. 2. MH02 EuRoC dataset sample bucket pairs which are extracted at t_1 and t_2, and their intensity differences histogram.

as the 1st order Markov model and propose the weighting matrix as a function of previously converged velocity. More specifically, the weighting matrix is calculated as follow,

$$W_P = \alpha ||\xi_P||_2 I_6, \tag{19}$$

where α is a constant slope and I_6 is 6 by 6 identity matrix.

We are motivated by the fact that the faster a camera moves, the more feasible it is affected by the previous pose because of its inertial force. Under a usual camera operation, the camera gathers images at a rate of 10-60 fps. Therefore, the time interval between incoming images is short enough to assume that the current estimator is highly influenced by how fast the previous pose was.

传入图像之间的时间间隔足够短，可以假设当前的估计量很大程度上受先前姿势的速度影响

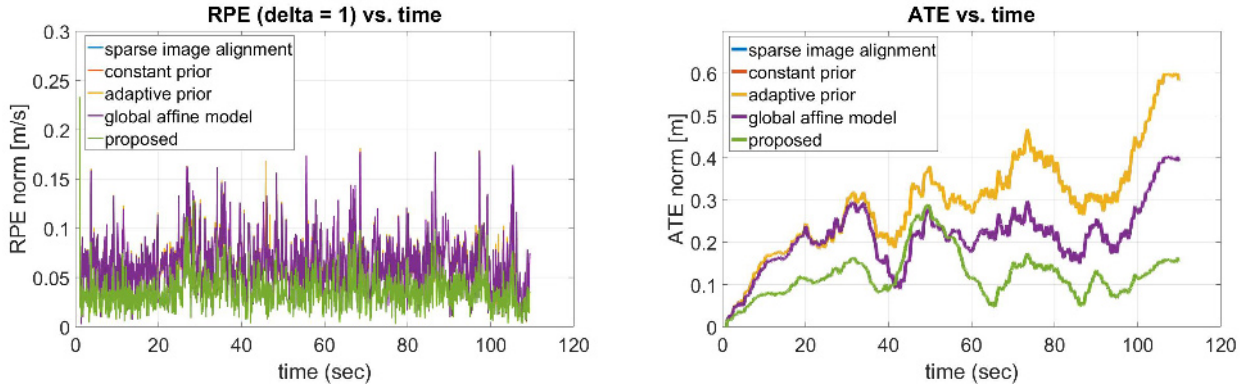
4. EXPERIMENTAL RESULTS

4.1. MAV dataset

The algorithm is evaluated at the real-world dataset, EuRoC dataset which is recorded by the stereo camera mounted at the MAV [15]. We have to mention that even if the image sequences are successive, there exist substantial illumination changes violating the photo-consistency assumption. Fig. 2(a) shows sample buckets extracted at the pair of temporally consecutive images with the interval of 50 ms. Specifically, A_i and B_i are buckets at the timestamp $t_i\,(i = 1, 2)$, for example, the pair of A_1, A_2 corresponds to the same bucket at the different instance. To verify illumination changes, we draw the intensity difference histogram for the pair of A_i and B_i in Fig. 2(b). We observe that intensity differences are not negligible, also the histograms for the bucket pairs are not identical to each other in Fig. 2(b). Therefore, to obtain reliable pose of the MAV, local and global illumination changes should be considered.

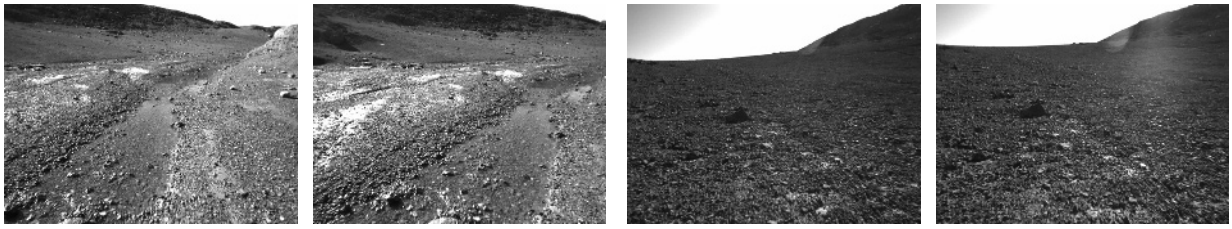
For implementation details, we obtain feature correspondences between the left and right stereo image using minimum eigenvalue feature detection [19] and Kanade Lucas Tomasi (KLT) tracker [20]. Note that the KLT tracker does not track features at temporal images but features at static stereo images where extrinsic parameter, ${}^{C_2}T_{C_1}$ is calibrated in advance. Also, we maintain 100-150 number of 3×3 patches in 5×5 buckets at 20 fps image sequence, and to suppress large photometric errors, we employ T-distribution image weighting matrix as in [17]. Lastly, we iteratively solve the optimization problem using the Levenberg-Marquardt algorithm, and the proposed algorithm is implemented in MATLAB.

The ground truth trajectory and attitude are provided by a motion capture system, and the MAV flies 63.2-meter long trajectory for 110 seconds. We compare five different cases, i.e., sparse image alignment (sia), 'sia' with constant temporal prior weights (constant prior), 'sia' with adaptive temporal prior weights (adaptive prior), 'sia' with global affine illumination model (global) and the proposed algorithm (proposed). Three error metrics used for evaluating performance of each case are the root mean square error (RMSE) of the relative pose error (RPE) where the step size is equal to one that measures the local accuracy of the given trajectory, RMSE of the absolute trajectory error (ATE) and the final position error divided by the whole length of the ground truth trajectory (% dt). RPE and ATE are proposed by [21] and broadly used for evaluating VO algorithms.



(a) L_2 norm of RPE versus flight time. (b) L_2 norm of ATE versus flight time.

Fig. 3. Pose estimation accuracy at EuRoC MH02 dataset, 'sia', 'constant prior' and 'adaptive prior' are almost identical.



(a) (b) (c) (d)

Fig. 4. ASRL sample images, temporally consecutive image pairs that exhibit large motion, only three-quarters of the image is overlapped because of the large motion (a, b), and partial illumination change due to the sunlight (c, d).

Table 1. Performance comparison at the EuRoC MH02 dataset.

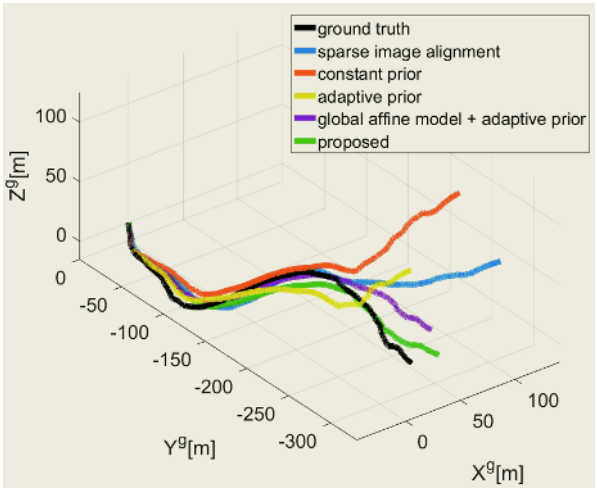
	RMSE RPE [m/s]	RMSE ATE [m]	% dt [%]
sia	0.0711	0.3213	0.93
constant prior	0.0710	0.3213	0.93
adaptive prior	0.0711	0.3212	0.93
global	0.0710	0.2256	0.62
proposed	0.0398	0.1350	0.25

The experimental results are summarized in Table 1. It reports that the proposed algorithm attains the most accurate pose estimation result. In particular, the proposed method has decreased RMSE RPE by 43.5%, RMSE ATE by 54.6%, and % dt by 58.8% on average of other methods in Table 1. We observe that 'adaptive prior' and 'constant prior' show almost the same results as 'sia' case. This is because the frame rate (20 fps) relative to the motion is high enough to prevent the state vector from falling into a false minimum. Fig. 3 shows the L_2 norm of RPE and ATE throughout the flight. At the pair of images in Fig. 2, the proposed algorithm shows 0.02 m/s of RPE norm whereas 'sia' shows 0.0308 m/s and 'global' shows 0.0312 m/s that is seen at 105.6 seconds elapsed time in Fig. 3(a). It is interesting to note that the proposed method further reduces the pose estimation error by modeling local brightness changes the global model fails to account for. As a result of the concatenation of relative poses, all five VO algorithms accumulates the ATE as in Fig. 3(b). However, the proposed algorithm has reduced the accumulation by considering partial brightness changes.

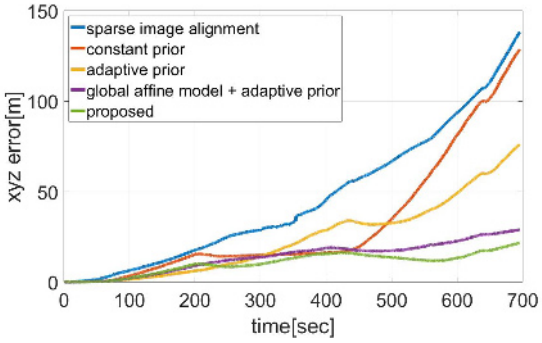
4.2. Planetary rover dataset

We evaluate the proposed algorithm at the planetary rover dataset, ASRL dataset. It is recorded by sensors equipped on the rover at Devon island located at Canadian High Arctic which exhibits strong geological terrains with no artificial objects and structures. Due to its diverse geological terrains without vegetation, it is utilized for planetary exploration field tests [14]. The dataset provides grayscale images at 3 fps with ground truth initial attitude and synchronized differential global positional system (DPGS) positions.

Fig. 4 shows sample images from the dataset featuring strong outdoor sunlight and large motion due to its low sampling time. More specifically, (a) and (b) in Fig. 4 are captured when the rover turned to the right, and only three quarters of the previous image, Fig. 4(a) remains overlapped with the current image, Fig. 4(b). Also, (c) and (d)



(a) Ground truth and estimated trajectory.



(b) xyz error versus time.

非常接近地面真值

Fig. 5. Pose estimation accuracy at ASRL dataset.

in Fig. 5 exhibit partial illumination changes occurred by the projection of the outdoor sunlight into the lens even though Fig. 4(c,d) are temporally successive. Therefore, local illumination changes and motion priors should be considered in order to accurately estimate rover's pose. The parameter settings are the same as MAV test except for bigger patch size (5×5) and fewer buckets (3×3). Also, note that we decide to employ Huber image weighting matrix [17] after trial and error to suppress large photometric errors. The ground truth and estimated trajectories are plotted in Fig. 5(a) and the rover drives 413-meter long trajectory for 11 minutes. We compare five algorithms as in the MAV dataset, and select error metrics as 3D position RMSE and % dt since the true attitude is not available in the dataset.

Table 2 summarizes the evaluation result that the proposed method outperforms the conventional methods; the proposed illumination model and the adaptive prior weights have lowered the position RMSE by 79.5 % with regards to 'sia' case. Fig. 5 shows the evaluation results and several interesting observations. First, the large motion of the rover degrades the accuracy of pose estimation making high nonlinearity to the cost function, (7). Thus, 'sia' case shows the largest position error accumulation as shown in Fig. 5, hence the worst position accuracy, 58.6 m as summarized in Table 2. The motion prior term in (8) holds the relative pose to stay near the previous motion stabilizing the estimator. Therefore, both 'constant prior' and 'adaptive prior' gives a more accurate estimation than 'sia' case. To reflect the importance of the prior term, we add the adaptive prior term to 'global' case to compare algorithms with the proposed one in fairness. Second, 'adaptive prior' shows 14.8 m better position accuracy than 'constant prior'. This is because we reflect the previous motion into the prior weighting matrix and the constant velocity residual is weighted accordingly. Third, even if the constant velocity model reduces the position error of the rover, both global and bucketed local illumination model reduce the error further. However, 'global' case cannot explain partial illumination changes such as image sequence in Fig. 4(c,d). We verify that the proposed algorithm is more robust to illumination changes than the global model through experimental evidence. The proposed algorithm yields the best position accuracy for the rover showing 4.2 m lower position RMSE than 'global' case. Lastly, a frame rate of a camera plays important role in DVO since the nonlinearity of the cost function is sensitive to how large overlapping region of a temporal image is.

Table 2. Performance comparison at ASRL dataset.

	Position RMSE [m]	% dt [%]
sia	58.6	33.5
constant prior	46.3	31.2
adaptive prior	31.5	18.4
global	16.2	7.01
proposed	12.0	5.25

5. CONCLUSION

In this paper, we have proposed the bucketed local illumination model and the adaptive prior weight in patch-based DVO framework. The proposed illumination model does not require depths for all pixels in the image while accounting for both global and local illumination changes. A further advantage of the proposed model is that it does not exploit artificial planar patches but small patches that are assumed to be planar patches in smooth surfaces. Fur-

本文在基于补丁的DVO框架中，提出了bucked局部光照模型和自适应先验权。在考虑全局和局部照明变化的同时，所提出的照明模型不需要图像中所有像素的深度。
该模型的另一个优点是，它不利用artificial planar，而是利用假定为光滑表面的planar patches。
此外，自适应先验权反映了这样一个事实：在统计上，快速运动的物体比慢速运动的物体对等速模型置信度更高。
实验结果表明，该算法对光照变化和大运动具有鲁棒性，比全局光照模型具有更好的姿态/位置精度

thermore, the adaptive prior weight reflects the fact that a fast-moving-object gives more confidence to the constant velocity model than a slow-moving-object statistically. Finally, we have evaluated our algorithm in the MAV and the planetary rover dataset where the camera's automatic exposure and the strong outdoor sunlight induce partial and sudden illumination changes. Our experimental result reports that the proposed algorithm is robust to illumination changes and large motions showing much better pose/position accuracy than the global illumination model.

REFERENCES

[1] S. Davide and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robotics & Automation Magazine*, vol. 18, no. 4, pp. 80-92, December 2011.

[2] C. Yang, M Maimone, and L. Matthies, "Visual odometry on the Mars exploration rovers," *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, pp. 903-910, October 2005.

[3] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: fast semi-direct monocular visual odometry," *Proc. of IEEE International Conference on Robotics & Automation*, pp. 15-22, May 2014.

[4] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, July 2004.

[5] F. Steinbrucker, J. Sturm, and D. Cremers, "Real-time visual odometry from dense RGB-D images," *Proc. of IEEE International Conference on Computer Vision Workshops*, pp. 719-722, November 2011.

[6] V. Q. Nhat and G. Lee, "Illumination invariant object tracking with adaptive sparse representation," *International Journal of Control, Automation and Systems*, vol. 12, no. 1, pp. 195-201, February 2014.

[7] P. Kim, B. Coltin, O. Alexandrov, and H. J. Kim, "Robust visual localization in changing lighting conditions," *Proc. of IEEE International Conference on Robotics and Automation*, pp. 5447-5252, June 2017.

[8] S. Klose, P. Heise, and A. Knoll, "Efficient compositional approaches for real-time robust direct visual odometry from RGB-D data," *Proc. of IEEE International Conference on Intelligent Robots and Systems*, pp. 1100-1106, November 2013.

[9] J. Jordan and A. Zell, "Ground plane based visual odometry for RGBD-cameras using orthogonal projection," *International Federation of Automatic Control on Intelligent Autonomous Vehicles*, pp. 108-113, June 2016.

[10] H. Jin, P. Favaro, and S. Soatto, "Real-time feature tracking and outlier rejection with changes in illumination," *Proc. of IEEE International Conference on Computer Vision*, pp. 684-689, July 2001.

[11] J. Engel, J. Stuckler, and D. Cremers, "Large-scale direct SLAM with stereo cameras," *Proc. of IEEE International Conference on Intelligent Robots and Systems*, pp. 1935-1942, September 2015.

[12] P. Kim, H. Lim, and H. J. Kim, "Robust visual odometry to irregular illumination changes with RGB-D camera," *Proc. of IEEE International Conference on Intelligent Robots and Systems*, pp. 3688-3694, October 2015.

[13] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," *IEEE Intelligent Vehicles Symposium*, pp. 486-492, June 2010.

[14] P. Furgale, P. Carle, J. Enright, and T. D. Barfoot, "The Devon island rover navigation dataset," *The International Journal of Robotics Research*, vol. 31, no. 6, pp. 707-713, April 2012.

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

[16] J. Engel, J. Sturm, and D. Cremers, "Semi-dense visual odometry for a monocular camera," *IEEE International Conference on Computer Vision*, pp. 1449-1456, December 2013.

[17] C. Kerl, J. Sturm, and D. Cremers, "Robust odometry estimation for RGB-D cameras," *Proc. of IEEE International Conference on Intelligent Robots and Systems*, pp. 3748-3754, May 2013.

[18] C. Kerl, *Odometry from RGB-D Cameras for Autonomous Quadcopters*, Master's Thesis, Technical University of Munich, 2012.

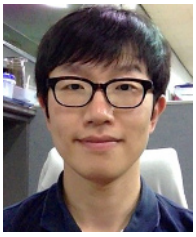
[19] J. Shi, "Good features to track," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 593-600, June 1994.

[20] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proceedings DARPA Image Understanding Workshop*, pp. 121-130, April 1981.

[21] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers "A benchmark for the evaluation of RGB-D SLAM systems," *Proc. of IEEE International Conference on Intelligent Robots and Systems*, pp. 573-580, October 2012.

[22] X. Zheng, Z. Moratto, M. Li, and A. I. Mourikis, "Photometric patch-based visual-inertial odometry," *Proc. of IEEE International Conference on Robotics and Automation*, pp. 3264-3271, May 2017.

Jae Hyung Jung is an M.S. student in the Department of Mechanical and Aerospace Engineering of Seoul National University, Korea. He received the B.S. degree in the Department of Aerospace Engineering from Pusan National University, Korea in 2017. His research interests include visual odometry and vision-aided inertial navigation for mobile robots.



Sejong Heo is a Ph.D. student in the Department of Mechanical and Aerospace Engineering of Seoul National University, Korea. He received the B.S. and M.S. degrees in the Department of Mechanical and Aerospace Engineering from Seoul National University, Korea, in 2008 and 2010, respectively. He worked for Doosan DST in Korea, which is the maker of the high precision INS. His current research topics include the high precision inertial navigation, Bayesian filtering, nonlinear optimization and vision-aided inertial navigation for land vehicles and mobile robots.

Chan Gook Park received the B.S., M.S., and Ph.D. in control and instrumentation engineering from Seoul National University, South, Korea, in 1985, 1987, and 1993, respectively. He worked with Prof. Jason L. Speyer on peak seeking control for formation flight at the University of California, Los Angeles (UCLA) as a postdoctoral fellow in 1998. From 1994 to 2003, he was with Kwangwoon University, Seoul, Korea, as an associate professor. In 2003, he joined the faculty of the School of Mechanical and Aerospace Engineering at Seoul National University, Korea, where he is currently a professor. From 2009 to 2010, he was a visiting scholar with the Department of Aerospace Engineering at Georgia Institute of Technology, Atlanta, GA. He served as a chair of IEEE AES Korea Chapter until 2009. His current research topics include advanced filtering techniques, high precision INS, GPS/INS integration, MEMS-based pedestrian dead reckoning, and visual inertial navigation.