Vehicle body full SE (3) pose estimation using Surround-view System

Manasvi Saxena

Mercedes Benz R&D India Pvt. Ltd., Bangalore, India

saxena.manasvi06@gmail.com

ABSTRACT

In this paper, I provide a Theoretical derivation for Vehicle body SE(3) pose estimation using Images obtained from 4 camera attached around the vehicle by minimizing the Photometric errors. The derivation here is inspired from the recent paper Xiao Liu, Lin Zhang et al. [1] but with a different strategy as explained below, which make this contribution novel.

In the paper of Xiao Liu, Lin Zhang et al. the main idea was to create a consistent surround view image from the well calibrated camera poses. The calibration is normally done offline which may not be sufficient as the camera poses can change when the vehicle experiences other dynamic motions while driving. Therefore, the authors of the paper developed finally a Ground-Camera Model which determines the correct SE(3) poses of the two (out of 4) cameras by minimizing the difference in the Intensity of the image pixels captured by the camera chosen for correction with the Intensity of the image pixels obtained from the ground projections of the two cameras adjacent to the one chosen for correction. The intensity difference is calculated in the image plane of the camera chosen for correction. This process is done individually for each of the two cameras starting with best initial known pose and later optimized until the photometric error reaches minimum. Using the optimum poses for the two cameras results in a consistent surround view image.

In contrast, I have again derived the Ground-Camera model equations with the SE(3) pose of a frame fixed to the vehicle rigid body as the only variable in optimization. This approach addresses directly the underlying reason of the main source behind the change of the camera poses which is in fact the change of the pose of Vehicle rigid body itself as the 4 cameras are rigidly mounted on the vehicle body. Therefore, I propose to estimate the optimum SE(3) pose of the Vehicle body that minimizes total photometric error for each of the adjacent camera pairs by calculating the difference in Intensity of the image pixels captured by one of the camera (in the pair) with the intensity of the image pixels obtained from the ground projections of the other camera (in the pair). The intensity difference is calculated in the image plane of the former. The projections are performed starting with the best initial known pose and later optimized until the photometric error reaches minimum.

INTRODUCTION

In this paper, I only explain the proposed approach briefly and show the derivation of the Ground-Camera model equations with the SE(3) pose of a frame fixed to the vehicle rigid body as the only variable in optimization. For a more detailed understanding of the Ground-Camera model, the reader is requested to refer the parent paper [1]. We further developed their approach and have again derived the equations to estimate the Vehicle's rigid body pose instead of individually estimating the pose of the two cameras which will though results in a consistent surround view image but doesn't addresses directly the underlying reason of the main source behind the change of pose of the cameras (i.e. change of pose of the Vehicle's rigid body).

THE PROPOSED APPROACH

In this section I will provide the details of the derivation of the Ground-Camera model equations with the SE(3) pose of a frame fixed to the vehicle rigid body as the only variable in optimization. Let us focus on the common overlap region from the Left & the Back Camera as shown in Fig. 1. The derivation for the other three common overlap regions i.e. Back & Right, Right & Front and Front & Left is similar.

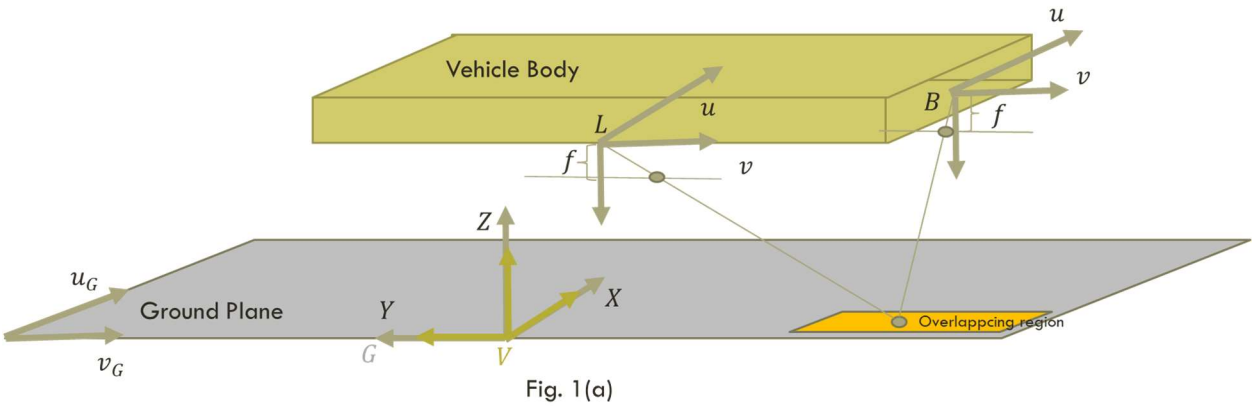

Fig. 1(a)

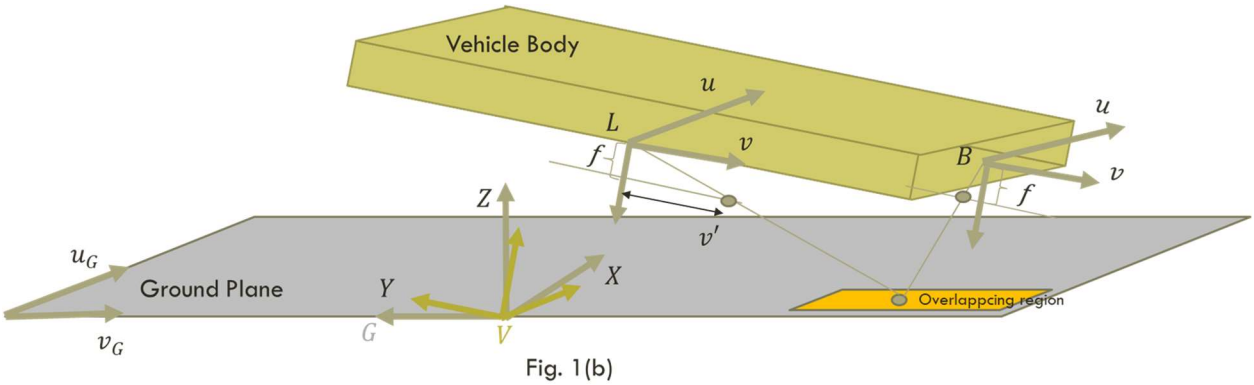

Fig. 1(b)

Figure 1 (a) shows the initial configuration of the Vehicle rigid body and a pair of adjacent camera Left (L) & Back (B) w.r.t Ground plane. The overlapping region which is commonly

observed by both the camera is highlighted in orange. Figure 1 (b) shows the configuration of the Vehicle rigid body and a pair of adjacent camera Left (L) & Back (B) w.r.t Ground plane after some dynamic motion experienced by the vehicle while driving. The SE(3) pose of the coordinate frame V which is fixed to the vehicle body but initially assumed to lie at the ground plane is the variable to be determined by minimizing the photometric error. The two cameras are facing downwards with L & B representing their camera centers, f the focal length, u & v are the two coordinate axis in the 2d image plane. The two coordinate axis of the Ground Image plane are represented as u_G and v_G. The reference frame is represented as G.

A basic thought exercise as illustrated using Fig. 1 (a), (b) and (c) will help to understand the Ground-Camera model in more detail. Starting with Fig. 1(b) let us imagine that while the vehicle experienced a pitch motion, the Left and the Back cameras of the vehicle captured image of the common overlapping region between the two cameras highlighted in orange. The actual camera poses of the Vehicle or of the Left & Back Camera when the image was captured is unknown and is basically the variable need to be determined in this paper given the two captured images. The best available knowledge of the pose of the Vehicle body, Left and Back Camera w.r.t original reference Ground frame G (as in Fig. 1(a)) is the one which is obtained from offline calibration and is overlaid on Fig. 1(b) with cyan color and is shown together in Fig. 1(c). Using the initial known pose of camera L, the image captured by camera L written as I_L is first projected to the Ground plane to get I_{GL} and then again to the image plane of camera B as per its initial known pose to get the final image as I_{GL}^B. In the next step we find the difference in the intensities of image I_{GL}^B as compared to that in the original image I_B for each pixel location. This process is shown for 1 pixel at location v' in the image plane of camera L with intensity gray for example, when projected to the image plane of camera B using initial known poses of camera L and B, is observed at location v'' in the image plane of camera B, which instead sees the intensity as cyan at location v'' in its original image. This intensity difference is defined as the cost in the objective function which needs to be minimized by varying the pose of the Vehicle body.

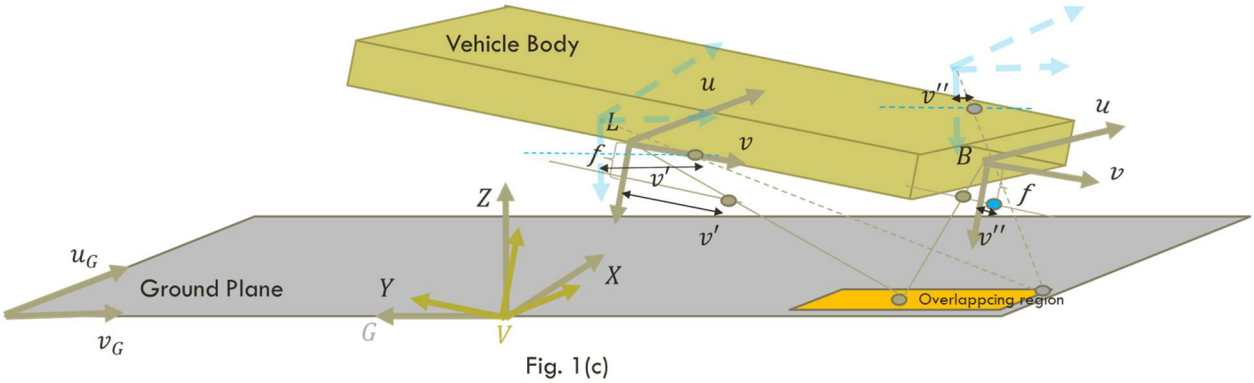

Fig. 1(c)

DERIVATION

As explained in previous section, the objective function with intensity difference as the cost is defined in eq. (1) for the pair of adjacent camera Left (L) & Back (B).

$$e_p = ||I_B(p) - I_{GL}^B(p)||_2 \tag{1}$$

Where, p is pixel location in the image plane of camera B and can be defined in eq. (2) in terms of camera intrinsic matrix K_B, and the position of the 3D point in the camera B frame. 3D points in the camera B frame or in the vehicle frame V can also be written as P_B and P_V like in eq. (3) & eq. (4) respectively in terms of pose of the vehicle frame V in camera B as T_{BV}, pose of the reference ground frame in the Vehicle frame V as T_{VG}, and 3d points in the reference ground frame as P_G. Please note that P_V is in the homogenous coordinates of a 3d vector $\overline{P_V}$. This will be used in eq. (12).

$$p = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_B} \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_B \\ Y_B \\ Z_B \\ 1 \end{bmatrix} = \frac{K_B P_B}{Z_B} \tag{2}$$

$$P_B = T_{BV} P_V = [X_B \quad Y_B \quad Z_B \quad 1]^T \tag{3}$$

$$P_V = T_{VG} P_G = [X_V \quad Y_V \quad Z_V \quad 1]^T = [\overline{P_V} \quad 1]^T \tag{4}$$

Thus, the objective function can be written as below in eq. (5) with small perturbation $\exp(\widehat{\xi_{VG}})$ applied on the manifold to the optimization variable which in my case is the pose of the reference ground frame in the Vehicle frame V, T_{VG}.

$$e_p = ||I_B(\frac{K_B T_{BV} \exp(\widehat{\xi_{VG}}) T_{VG} P_G}{Z_B}) - I_{GL}^B(\frac{K_B T_{BV} \exp(\widehat{\xi_{VG}}) T_{VG} P_G}{Z_B})||_2 \tag{5}$$

Now, the final step is to find the Jacobians of the objective function w.r.t variation in the tangent vector space of the Lie group i.e. ξ_{VG} to iteratively update the pose T_{VG} until the photometric error is minimized, as described in the parent paper [1]. The Jacobian can be decomposed as in eq. (6) where each term is explained in the following equations. In eq. (12) we use the property of the operator $\odot$ to write $\widehat{\xi_{VG}} P_V = P_v^{\odot} \xi_{VG}$ which then makes derivative straight forward as defined in Barfoot [2, Ch. 7].

$$J = \frac{\partial e_p}{\partial I_B} \frac{\partial I_B}{\partial p} \frac{\partial p}{\partial P_B} \frac{\partial P_B}{\partial \xi_{VG}} + \frac{\partial e_p}{\partial I_{GL}^B} \frac{\partial I_{GL}^B}{\partial p} \frac{\partial p}{\partial P_B} \frac{\partial P_B}{\partial \xi_{VG}} \tag{6}$$

$$\frac{\partial e_p}{\partial I_B} = I_B(p) - I_{GL}^B(p) \tag{7}$$

$$\frac{\partial e_p}{\partial I_{GL}^B} = -(I_B(p) - I_{GL}^B(p)) \tag{8}$$

$$\frac{\partial I_B}{\partial p} = [\nabla u_B \quad \nabla v_B \quad 0] \tag{9}$$

$$\frac{\partial I_{GL}^{B}}{\partial p} = [\nabla u_{GL}^{B} \quad \nabla v_{GL}^{B} \quad 0] \tag{10}$$

$$\frac{\partial p}{\partial P_B} = \begin{bmatrix} \frac{f_x}{Z_B} & 0 & -\frac{f_x X_B}{Z_B^2} & 0 \\ 0 & \frac{f_y}{Z_B} & -\frac{f_y Y_B}{Z_B^2} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{11}$$

$$\frac{\partial P_B}{\partial \xi_{VG}} = \frac{\partial T_{BV} \exp(\widehat{\xi_{VG}})\, T_{VG} P_G}{\partial \xi_{VG}} = \frac{\partial T_{BV}\,(\mathrm{I} + \widehat{\xi_{VG}})\, T_{VG} P_G}{\partial \xi_{VG}} \tag{12}$$

$$= \frac{\partial T_{BV}\, \widehat{\xi_{VG}}\, P_V}{\partial \xi_{VG}} = \frac{\partial T_{BV}\, P_v^{\odot}\, \xi_{VG}}{\partial \xi_{VG}} = T_{BV}\, P_v^{\odot} = \begin{bmatrix} R_{BV} & t_{BV} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & -\widehat{P_V} \\ 0 & 0 \end{bmatrix}$$

REFERENCES

1. Xiao Liu, Lin Zhang et al. Online Camera Pose Optimization for the Surround-view system, 383-391, MM '19: Proceedings of the 27th ACM International Conference on Multimedia
2. T.D. Barfoot, State Estimation for Robotics. Cambridge University Press, 2017.