Visual Mapping and Pose Estimation for a Self-Driving Car with a Multi-Camera System

CVPR 2014 Doctoral Consortium

Gim Hee Lee (gimhee.lee@merl.com)
Mitsubishi Electric Research Labs, USA

Advisor: **Marc Pollefeys** (marc.pollefeys@inf .ethz.ch)
ETH Zürich, Switzerland
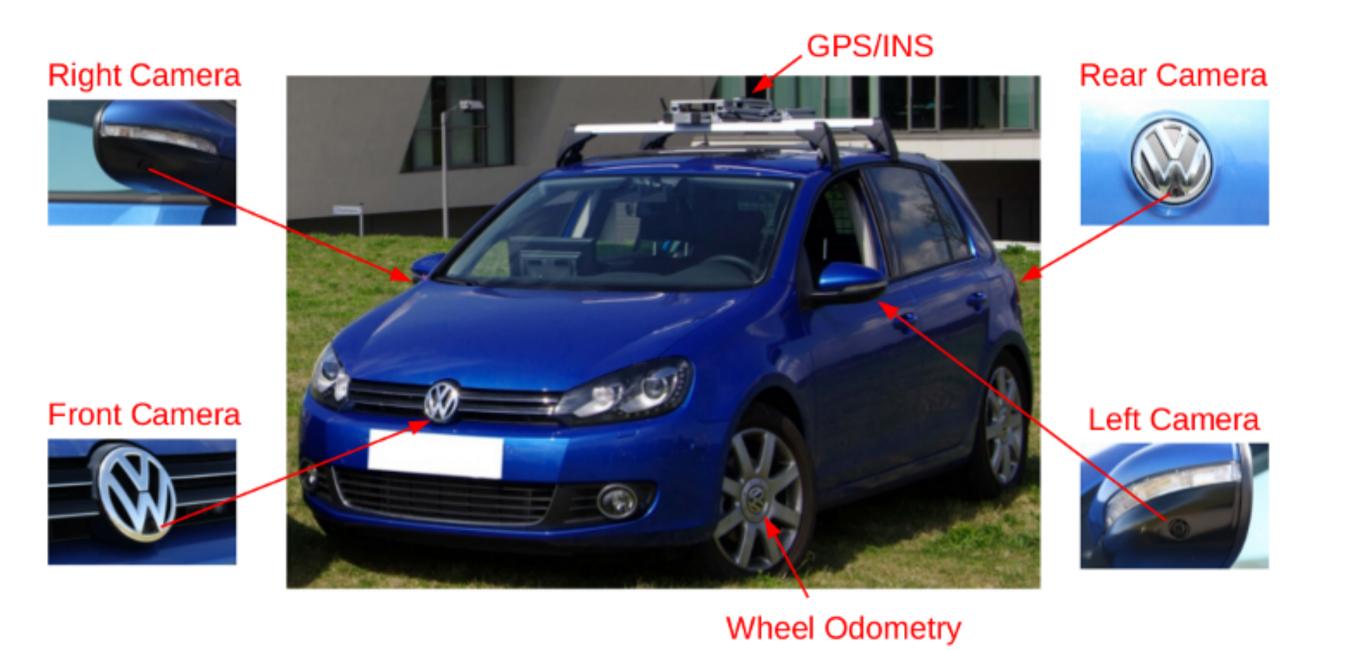
Introduction

- **Objectives:** Use images from a calibrated multi-camera system, i.e. known intrinsics and extrinsics parameters mounted on a self-driving car to:

  1. Generate a map, which we called the sparse map, that consists of the **ego-motions** of the car and 3D points from the structures in the environment.

  2. Estimate the pose of the car with respect to a given sparse map, which is also known as **non-perspective pose estimation**.

- **Grobi** - Our self-driving car equipped with wheel odometer, GPS/INS for ground truth, and four fisheye cameras looking front, left, right and rear.

Ego-Motion Estimation

- **Generalized Camera Model**

  Relative pose R and t of a multi-camera system can be solved from the **Generalized Epipolar Constraint** defined in [1]:
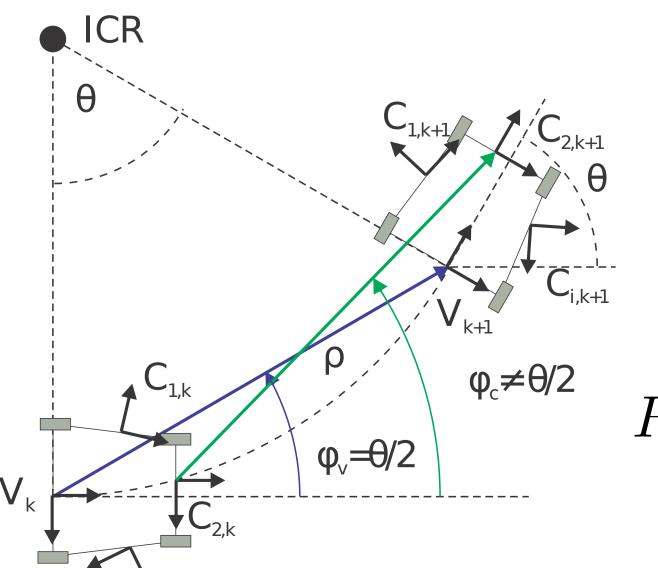
$$l'^{T}_{ij} \underbrace{\begin{bmatrix} E & R \\ R & 0 \end{bmatrix}}_{E_{GC}} l_{ij} = 0 \qquad (1)$$

  where E_{GC} is the generalized essential matrix. $l \leftrightarrow l'$ are point correspondences represented as 6-vector Plücker lines. $E = \lfloor t \rfloor_\times R$ is the conventional essential matrix.

- **Existing Works and Their Limitations**

  Linear algorithm [1] requires 17-point correspondences and minimal problem [2] gives 64 solutions, which are computationally expensive for RANSAC.

- **Our Solution: Ackermann Constraint [3]**

  - Two parameters - relative yaw angle θ and scale $\rho \rightarrow$ 2-point minimal problem.

$$R = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad t = \rho \begin{bmatrix} \cos\varphi_v \\ \sin\varphi_v \\ 0 \end{bmatrix} \quad (2)$$

- Putting Eq (2) into E_{GC} from Eq (1), we get the generalized essential matrix with Ackermann motion.

$$E_{GC} = \begin{bmatrix} 0 & 0 & \rho\sin\frac{\theta}{2} & \cos\theta & -\sin\theta & 0 \\ 0 & 0 & -\rho\cos\frac{\theta}{2} & \sin\theta & \cos\theta & 0 \\ \rho\sin\frac{\theta}{2} & \rho\cos\frac{\theta}{2} & 0 & 0 & 0 & 1 \\ \cos\theta & -\sin\theta & 0 & 0 & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (3)$$

- Putting Eq (3) into Eq (1) and with two Plücker line correspondences, we get a system of two polynomial equations in the form of:

$$a\cos\theta + b\sin\theta + c\rho\cos\frac{\theta}{2} + d\rho\sin\frac{\theta}{2} + e = 0 \quad (4)$$

  a, b, c, d and e are coefficients computed from the Plücker line.

- This leads to the minimal solution of 3-degree polynomial with $\gamma = \sin^2\frac{\theta}{2}$ where the yaw angle θ can be solved in closed-form.
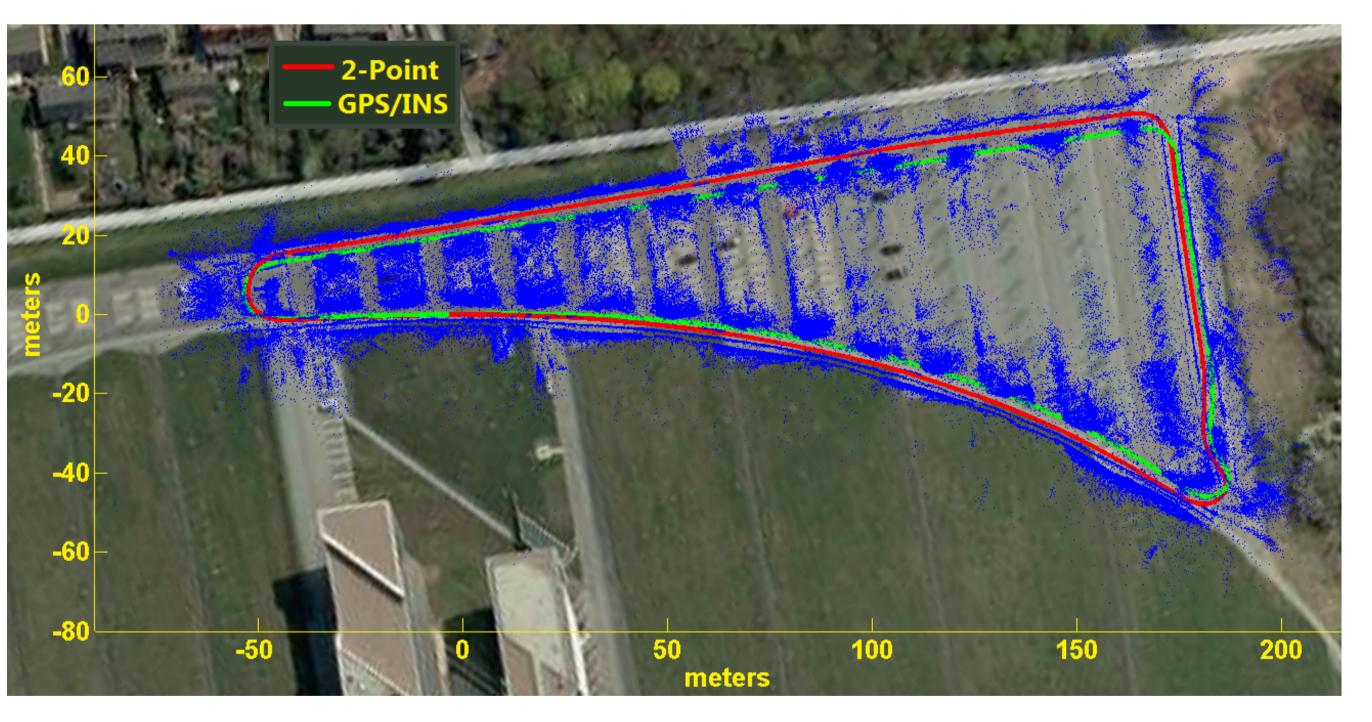
$$A\gamma^3 + B\gamma^2 + C\gamma + D = 0$$

  A, B, C and D are the coefficients computed from the Plücker line correspondence $l \leftrightarrow l'$.

- The metric scale ρ can be solved by back-substitution.

- **Results**

  Top view of trajectory and 3D map points after pose-graph loop-closure and full bundle adjustment compared with GPS/INS ground truth.

Loop-Closure Constraint

- Errors from ego-motion estimation accumulate and eventually lead to loop-closure errors.

- Figure shows an example of loop-closure error. Pose-graph (red) does not coincide at revisited places.

- **Problem:** Loop-closure constraints (green) do not follow Ackermann constraint.

- **Solution:** Relax Ackermann to planar constraint, i.e. three parameters - $x, y, \theta \rightarrow$ 3-point minimal problem [4].

- Planar constraint:

$$R = \frac{1}{1+q^2}\begin{bmatrix} 1-q^2 & -2q & 0 \\ 2q & 1-q^2 & 0 \\ 0 & 0 & 1+q^2 \end{bmatrix}, \quad t = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \quad (5)$$
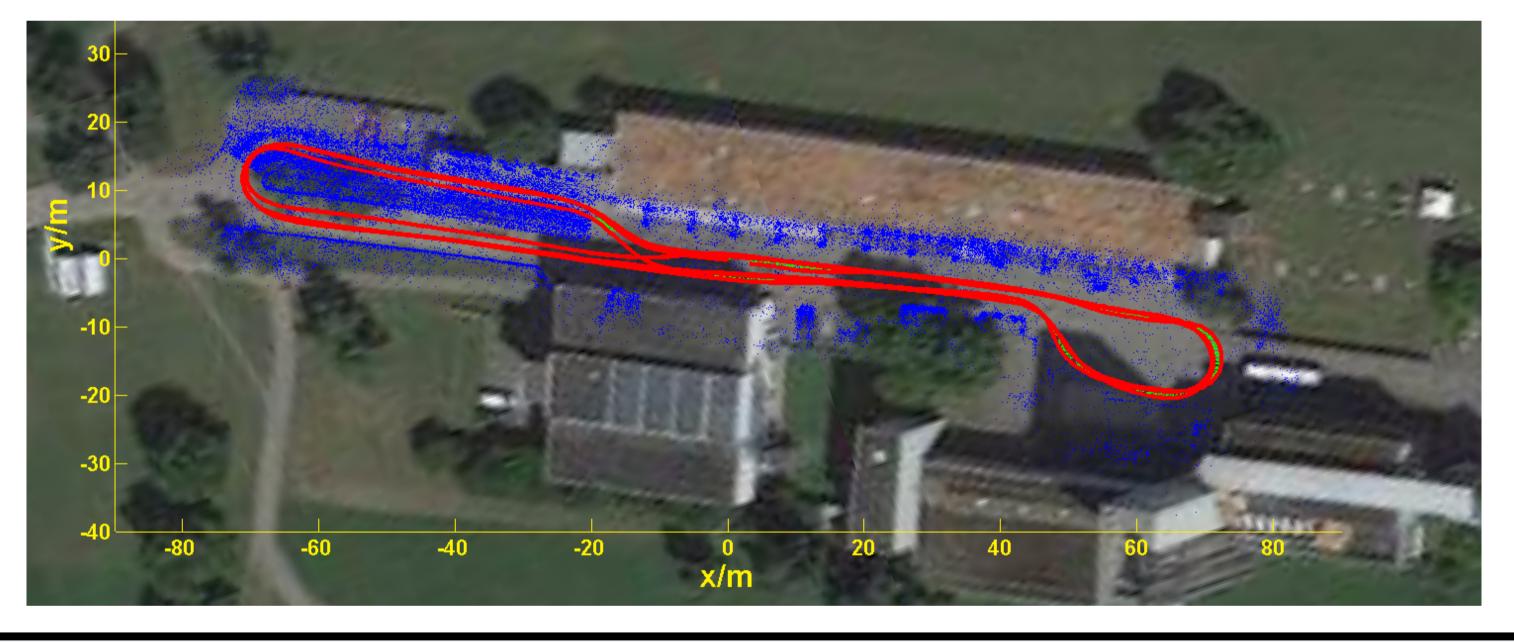
  where $q = tan(\frac{\theta}{2})$, hence $cos(\theta) = \frac{1-q^2}{1+q^2}$ and $sin(\theta) = \frac{2q}{1+q^2}$.

- Putting Eq (5) into Eq (1) and with three Plücker correspondences, we get a system of three polynomial equations in the form of:
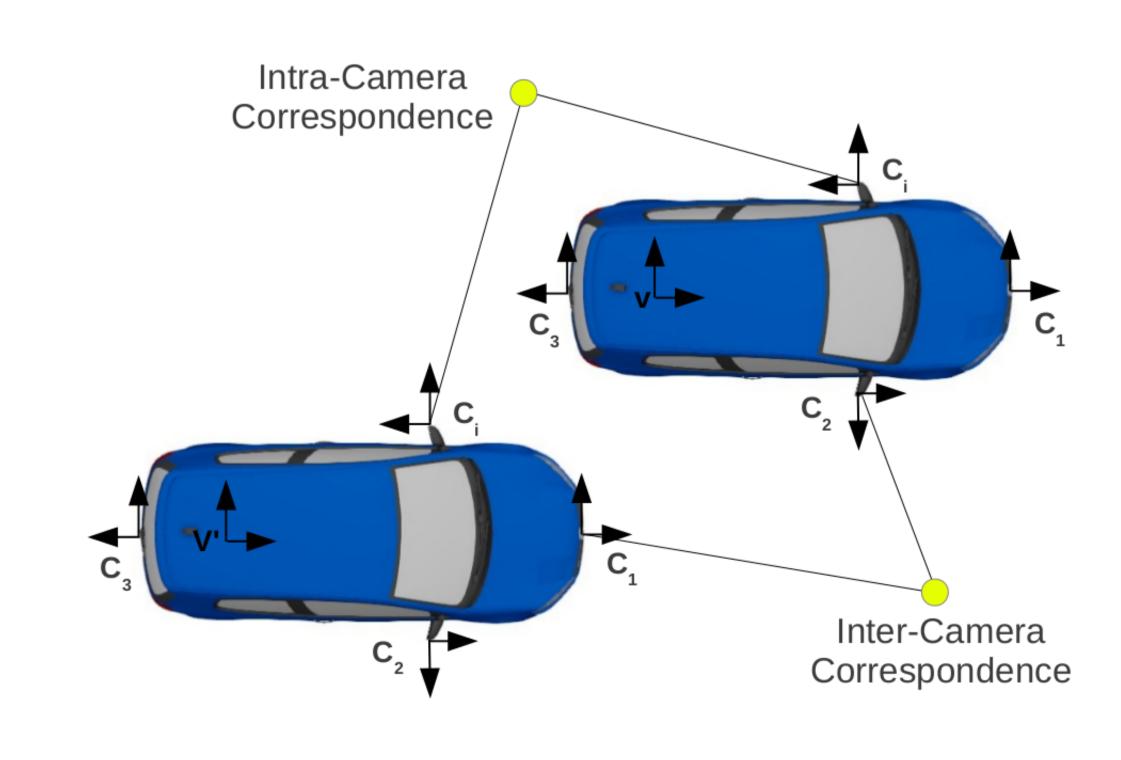
$$a_1 x q^2 + a_2 xq + a_3 x + a_4 yq^2 + a_5 yq + a_6 y + a_7 q^2 + a_8 q + a_9 = 0 \quad (6)$$

- The system of polynomial equations is solved with the "Hidden Variable Resultant" method.

- This gives a 6-degree polynomial that can be solved with the Companion matrix.

- **Results**

  Pose-graph (red) and 3D scene points (blue) after pose-graph optimization overlaid on the satellite image.

Degenerate Case and Its Solution

- Scale cannot be estimated from the 2-point and 3-point algorithms when the multi-camera system undergoes pure translation while having only intra-camera correspondences.

- We use an additional inter-camera correspondence to circumvent the scale problem in the degenerate case for the ego-motion estimation.

- We drop the loop-closure constraints in the rare cases of pure translational alignment during loop-closure.

Non-Perspective Pose Estimation

- **Objective**

  Given: A set of three 3D points (X_1, X_2, X_3) defined in F_W seen by arbitrary cameras on the multi-camera system and their corresponding 2D image coordinates (x_1, x_2, x_3).

  Find: The rigid transformation R and t that brings the multi-camera frame F_G into the world frame F_W.

- 3D points are defined as the Plücker coordinates:

$$X_i^G = q_i \times q_i' + \lambda_i q_i \quad (7)$$

  where λ_i is the depth of the point X_i^G along the Plücker line

- Using the preservation of distances between the 3D points in F_W and F_G, we get three constraints in the three unknown depth $(\lambda_1, \lambda_2, \lambda_3)$ [5]:

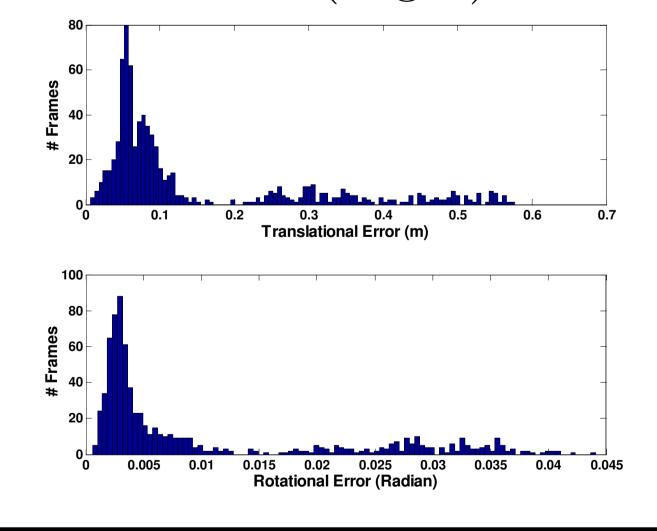$$k_{11}\lambda_1^2 + (k_{12}\lambda_2 + k_{13})\lambda_1 + (k_{14}\lambda_2^2 + k_{15}\lambda_2 + k_{16}) = 0 \quad (8a)$$

$$k_{21}\lambda_1^2 + (k_{22}\lambda_3 + k_{23})\lambda_1 + (k_{24}\lambda_3^2 + k_{25}\lambda_3 + k_{26}) = 0 \quad (8b)$$

$$k_{31}\lambda_2^2 + (k_{32}\lambda_3 + k_{33})\lambda_2 + (k_{34}\lambda_3^2 + k_{35}\lambda_3 + k_{36}) = 0 \quad (8c)$$

- The system of polynomial equations is solved using the "Sylvester Resultant" variable elimination method.

- This gives a 8-degree polynomial that can be solved with the Companion matrix.

- **Results:**

  Localization results (Left). Results from frames with < 20 correspondences are discarded. Plots showing the distribution of the translational and rotational errors against GPS/INS ground truths (Right).

References

1. R. Pless, "Using many cameras as one", CVPR 2003 .
2. H. Stewénius *et. al*, "Solutions to minimal generalized relative pose problems", OMNIVIS 2005.
3. G.H. Lee *et. al*, "Motion estimation for a self-driving car with a generalized camera", CVPR 2013.
4. G.H. Lee *et. al*, "Structureless pose-graph loop-closure with a multi-camera system on a self-driving car", IROS 2013.
5. G.H. Lee *et. al*, " Minimal solutions for pose estimation of a multi-camera system", ISRR 2013.