

Camera platforms for localization and map building

Henrik Stewenius, Magnus Oskarsson and Kalle Åström
Centre for Mathematical Sciences, LTH.
{stewe, magnuso, kalle}@maths.lth.se

Abstract

Vision is useful for the autonomous navigation of vehicles. In this paper the case of a vehicle equipped with multiple cameras with non-overlapping views is considered. The geometry and algebra of such a moving platform of cameras are considered. In particular we formulate and solve structure and motion problems for a few novel cases of such moving platforms. For the case of two-dimensional retina cameras (ordinary cameras) there are two minimal cases of three points in two platform positions and two points in three platform positions. In the paper it is also discussed how classical algorithms such as intersection and resection can be extended to this new situation. We also present a system for feature detection, tracking and structure and motion estimation for such platforms. The theory has been tested on synthetic and real data with promising results.

1 Introduction

Vision is useful for the autonomous navigation of vehicles. We will in this paper address the problem of estimating the position of the vehicle as well as the positions of world points using only image measurements. This is known as the structure and motion problem in computer vision. In robotics this problem is known as simultaneous localization and mapping, **SLAM**, or concurrent mapping and localization, **CML**. Most SLAM systems for robot navigation are based on sonar or laser range finders, but there are systems based on vision, cf. [2], and systems that use bearings, cf. [3]. For an overview of the SLAM problem see [7].

In this paper we consider a platform (a vehicle, robot, car) moving with a planar motion. This vehicle has a number of cameras with different camera centres facing outwards so that they cover different viewing angles. The purpose here is to get a large combined field of view using simple and cheap cameras. The cameras are assumed to have different camera centres but it is assumed that the cameras are calibrated relative to the vehicle. In [5] a similar setup was considered.

2 System overview

Consider m cameras that are fixed to a vehicle and consider images taken by these cameras at n different times, where the vehicle has moved. Assume that a number of corresponding points (possibly in different images) are measured. Then there are a number of problems that are interesting to look at.

1. **Calibration.** Assume that each camera sees enough points to calculate its own relative camera motion. Since the relative motion of all cameras is the same, how many cameras and views are needed to solve for the common motion. When is it possible to solve for the cameras' relative positions?
2. **Structure and motion.** Assume that the camera's position relative to the vehicle is known, but not the vehicle's motion, represented by a transformation matrix \mathbf{S}_i . Given a number of corresponding points, how should one calculate the world points \mathbf{U}_j and the transformations \mathbf{S}_i from image data?
3. **Intersection.** Assume that the camera's position relative to the vehicle is known, and that also the vehicle's motion is known. Assume that a number of corresponding points are measured, how should one calculate the world points \mathbf{U}_j from image data?
4. **Resection.** Assume that the camera's position relative to the vehicle is known, but not the vehicle's motion, represented by a transformation matrix \mathbf{S}_i . Assume that a number of corresponding points (possibly in different images) are measured, how should one calculate the transformations \mathbf{S}_i from image data and known world points \mathbf{U}_j ?

In this paper we assume that the calibration of the cameras relative to the vehicle has been done. In the experiments in section 5 the calibration was done manually. One could also consider autocalibration approaches similar to the approach in [1], where the problem of

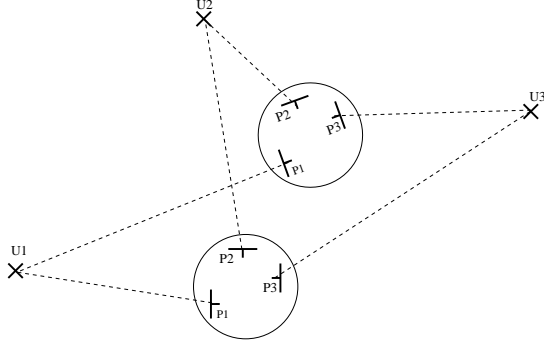


Figure 1: Three calibrated cameras with constant and known relative positions taking two images each

aligning video sequences was addressed, or the calibration problems in robotics [8] which are similar in nature to the calibration of cameras relative a vehicle.

The basic outline of our system is given in algorithm 2.1.

Algorithm 2.1 SLAM

1. Track features.
2. Use RANSAC and minimal data to find initial estimates.
3. Extend solution using intersection and resection.
4. Optimize solution using bundle adjustment.

We start by tracking features in the images. In our experiments we have used points. We then use the minimal cases described in section 3 to bootstrap the tracking and structure and motion estimation in a RANSAC [4] approach. One problem with RANSAC estimation is that one uses a minimal amount of data to estimate the model, which can lead to low accuracy. Since our vehicle is equipped with multiple cameras with a large combined field of view, we can get quite good initial estimates, see section 5. After an initial estimate has been found this is then extended and optimized using the techniques described in section 4.

3 Geometry and initial estimates

The standard pinhole camera model is used $\lambda \mathbf{u} = \mathbf{P}\mathbf{U}$, where the camera matrix \mathbf{P} is a 3×4 matrix. A scene point \mathbf{U} is in \mathbb{P}^3 and a measured image point \mathbf{u} is in \mathbb{P}^2 . It is sometimes useful to consider dual image coordinates. Each image point \mathbf{u} is dual to two linearly independent dual vectors \mathbf{v}^1 and \mathbf{v}^2 with $\mathbf{v}^i \mathbf{u} = 0$. The measurement equation then becomes

$$\mathbf{v}^1 \mathbf{P}\mathbf{U} = \mathbf{v}^2 \mathbf{P}\mathbf{U} = 0. \quad (1)$$

As the platform moves the cameras move together. This is modeled as a transformation \mathbf{S}_i between the first position and position i . In the original coordinate system the camera matrix for camera k at position i is $\mathbf{P}_k \mathbf{S}_i$.

It is assumed here that the camera views do not necessarily have common points. In other words, a point is typically seen by only one camera. On the other hand it can be assumed that in a couple of neighboring frames a point can be seen in the same camera. Assume here that point j is visible in camera k . The measurement equation for the n points is then

$$\lambda_{ij} \mathbf{u}_{ij} = \mathbf{P}_k \mathbf{S}_i \mathbf{U}_j, \quad j = 1, \dots, n, i = 1, \dots, m. \quad (2)$$

Note that $\mathbf{l}_{ij}^k T = \mathbf{v}_{ij}^k \mathbf{P}_j$ corresponds to the viewing plane in the vehicle coordinate system. Thus the constraint, in dual form, can be written

$$\begin{cases} \mathbf{l}_{ij}^1 \mathbf{S}_i \mathbf{U}_j = 0, \\ \mathbf{l}_{ij}^2 \mathbf{S}_i \mathbf{U}_j = 0, \end{cases} \quad j = 1, \dots, n, i = 1, \dots, m. \quad (3)$$

Here the planes \mathbf{l} are measured. The question is if one can calculate structure \mathbf{U}_j and motion \mathbf{S}_i from these measurements. Based on the previous sections, the structure and motion problem will now be defined.

Problem 1 Given n image points from m different platform positions \mathbf{u}_{ij} , $i = 1, \dots, m, j = 1, \dots, n$, and the camera matrices \mathbf{P}_j , $j = 1, \dots, n$, the **structure and motion problem** is to find reconstructed points, \mathbf{U}_j , and platform transformations, \mathbf{S}_i :

$$\mathbf{U}_j = \begin{pmatrix} X_j \\ Y_j \\ Z_j \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{S}_i = \begin{pmatrix} a_i & b_i & 0 & c_i \\ -b_i & a_i & 0 & d_i \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (4)$$

$$a_i^2 + b_i^2 - 1 = 0$$

such that

$$\lambda_{ij} \mathbf{u}_{ij} = \mathbf{P}_j \mathbf{S}_i \mathbf{U}_j, \quad \forall i = 1, \dots, m, j = 1, \dots, n$$

for some λ_{ij} .

In order to understand how much information is needed to solve the structure and motion problem, one can calculate the number of degrees of freedom of the problem and the number of constraints given by the projection equation. Each object point has three degrees of freedom. Vehicle location for a planarily moving vehicle has three degrees of freedom when using $a_i^2 + b_i^2 = 1$, that is Euclidean information and four degrees of freedom in the similarity case when not using this information. The word ‘‘image’’ is used to mean ‘‘All the information collected by all our cameras at one instant in time’’.

Table 1: The number of excess constraints $2mn - (3n + 3(m - 1))$ for the structure and motion problem with m images of n points.

n	m				
	1	2	3	4	5
1	-1	-2	-3	-4	-5
2	-2	-1	0	1	2
3	-3	0	3	6	9
4	-4	1	6	11	16
5	-5	2	9	16	23

For Euclidean reconstruction there are $2mn - (3n + 3(m - 1))$ excess constraints and as seen in table 1 there are two interesting cases, namely “Two images and three points ($m=2, n=3$)” and “Three images and two points ($m=3, n=1$)”.

These two problems can be solved and both cases lead to multiple solutions. The proofs are omitted due to lack of space.

Theorem 3.1 *For three calibrated cameras that are rigidly fixed relative to each other, each taking an image of a point at two distinct times there generally exist one or three real non-trivial solutions.*

Theorem 3.2 *Two cameras mounted rigidly with respect to each other for which calibration as well as relative positions are known are moved planarily to 3 different stations where they observe one point per camera. Under these circumstances there generally exist one or three non-trivial real solutions.*

For both cases pure translation is a degenerate case and can only be computed up to scale. The solution for the two problems leads to polynomials of low degree which can be solved quickly, robustly and accurately.

4 Refinement of solution

Intersection

Generalization of the intersection algorithm [4] to this new situation is straightforward. When both calibration $\mathbf{P}_1, \dots, \mathbf{P}_K$ and platform motion $\mathbf{S}_1, \dots, \mathbf{S}_m$ is known it is straightforward to calculate scene points coordinates \mathbf{U} from image measurements by intersecting view-lines. A linear initial solutions is obtained by solving

$$\begin{pmatrix} \mathbf{u}_1 \times \mathbf{P}_k \mathbf{S}_1 \\ \dots \\ \mathbf{u}_n \times \mathbf{P}_k \mathbf{S}_n \end{pmatrix} \mathbf{U} = 0 \quad (5)$$

in a least squares sense. This solution can then be refined by non-linear optimization, cf. 4.

Resection

Generalization of the resection algorithm [4] is slightly more complex since the view lines do not go through a common points (the focal point) as in the ordinary camera resection algorithm.

Here we introduce a direct method for finding an initial estimate to the resection problem based on a linearized reprojection error. The idea is to solve $\mathbf{u}_j \times \mathbf{P}_k \mathbf{S} \mathbf{U}_j = 0$, $j = 1, \dots, n$ which is linear in \mathbf{S} . In our case there are non-linear constraints on \mathbf{S} , so we use the theory for constrained optimization to find the optimal solution under the constraints.

Parameterize the platform motion as in (4), using parameters $x = (a, b, c, d)$. The constraints $\mathbf{u}_j \times \mathbf{P}_k \mathbf{S} \mathbf{U}_j = 0$ is linear in these parameters. The linear constraints can be rewritten $f = Mx = 0$. However there is a non-linear constraint $g = a^2 + b^2 - 1 = 0$. Initial solution to the resection problem can be found by solving

$$\min_{xg=0} \sum_j |\mathbf{u}_j \times \mathbf{P}_k \mathbf{S}(x) \mathbf{U}_j|^2. \quad (6)$$

Introduce the Lagrangian $L(x, \lambda) = |Mx|^2 + \lambda g$. The solution to (6) can be found by $\nabla L = 0$. Here

$$\nabla_x L = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} + \lambda \begin{bmatrix} 2a \\ 2b \\ 0 \\ 0 \end{bmatrix} = 0. \quad (7)$$

Here one may solve for (c, d) and insert this in the above equation giving

$$(A - BD^{-1}C + 2\lambda I) \begin{bmatrix} a \\ b \end{bmatrix} = 0. \quad (8)$$

Here there is a non-trivial solution to (a, b) if and only if 2λ is one of the two eigenvalues of $A - BD^{-1}C$. For these two solutions (a, b) is determined up to scale. The scale can be fixed by $a^2 + b^2 = 1$. Finally (c, d) can be found from (8). This gives a reasonably good initial estimate on the resection parameters. This estimate is then improved by minimizing reprojection errors. Experience shows that only a few iterations are needed here.

Bundle adjustment

The discussion in sections 3 and 4 focused on finding initial estimates of structure and motion, and it is necessary to refine these estimates using non-linear optimization or bundle adjustment, cf. [6, 4]. The generalization of bundle adjustment to platform motions is straightforward. One wants to optimize platform motions \mathbf{S}_i and scene points \mathbf{U}_j so that reprojection error is minimized. The fact that the platform has a very large field of view makes bundle adjustment much better conditioned than what is common.

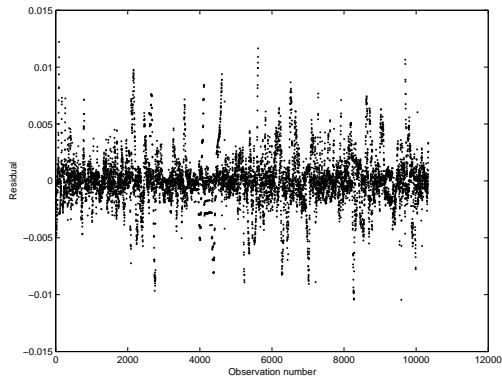


Figure 2: Residuals in the reprojected images.

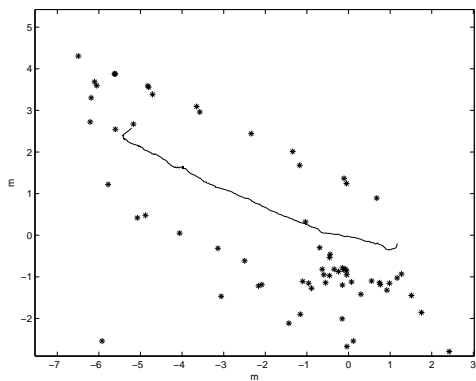


Figure 3: Reconstruction seen from above.

5 Experimental verification

A system with three video cameras was assembled and moved along a corridor with markers on the walls. Tracking was done by following these markers and some additional manual tracking.

As stated in section 3 pure translation is a degenerate case for the initial solver in that it will give a solution which is unknown up to scale. However some turns in the path will fixate the scale.

A reconstruction is shown in figure 3. The structure point which we seem to be passing through is a structure point which has a very high variance, as its measured view lines are almost collinear with the locations from which it is observed. The residuals in reprojection for this reconstruction are shown in 2. The size of the residuals are in the order of the error in a calibrated camera.

6 Conclusions

In this paper we have introduced the structure and motion problem for the notion of a platform of moving cameras. This problem is a generalization of the stan-

dard structure and motion problem in the sense that each time instant gives image measurements that correspond to view lines that not necessarily goes through the same point whereas in the standard structure and motion problem has view-lines that all go through the focal point. Two minimal cases for structure and motion estimation are solved, (i) three points in two views and (ii) two points in three views. Solution to the new type of resection problem is also given. Generalizations of the intersection algorithm and bundle adjustment are straightforward. Using these new solutions a system for feature detection, tracking, initial estimation, extension and bundle adjustment is presented.

Solutions to the minimal problems are useful for structure and motion estimation of autonomous vehicles equipped with multiple cameras. The existence of a fast solver for the two images and three points case is of interest when computing RANSAC. It is important to note that pure translation is a degenerate case and that the solution in this case suffers from the same unknown scale as for single camera solutions. Another important aspect is that the cameras have to have separate focal points.

In this paper the problem of calibrating the relative positions of the cameras on the platform is not studied. This is topic for further studies.

References

- [1] Y. Caspi and M. Irani. Alignment of Non-Overlapping sequences. In *Proc. 8th Int. Conf. on Computer Vision, Vancouver, Canada*, pages 76–83, 2001.
- [2] A. J. Davison and N. Kita. Sequential localisation and map-building in computer vision and robotics. In *SMILE 2000 Workshop, Dublin, Ireland*, pages 218–234, 2000.
- [3] M. C. Deans and M. Hebert. Experimental comparison of techniques for localization and mapping using a bearing-only sensor. In *7th intl. symp. on experimental robotics, Hawaii, USA*, 2000.
- [4] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [5] R. Pless. Using many cameras as one. In *Proc. Conf. Computer Vision and Pattern Recognition, Madison, USA*, 2003.
- [6] C. C. Slama. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA, 1980.
- [7] S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002.
- [8] H. Zhuang, Z. Roth, and R. Sudhakar. Simultaneous robot/world and tool/flange calibration by solving homogeneous transformations of the form $ax=yb$. *IEEE Trans. on Robotics and Automation*, 10(4):549–554, 1994.