Camera Calibration and Relative Pose Estimation from Gravity

Peter Sturm and Long Quan
INRIA Rhône-Alpes
655 Avenue de l'Europe
38330 Montbonnot St Martin, France
Peter.Sturm@inrialpes.fr

Abstract

We examine the potential use of gravity for camera calibration and pose estimation purposes. Concretely, objects being launched or dropped follow trajectories dictated by the law of gravity. We examine if video sequences of such trajectories give us exploitable constraints for estimating the imaging geometry. It is shown that it is possible to estimate the infinite homography and the epipolar geometry between pairs of views from this input, from which we can estimate (some) intrinsic parameters and relative pose. There are less singularities compared to approaches that do not use the information that the observed trajectories follow the gravity. In this paper, we sketch the geometric principles of our idea and validate them by numerical simulation.

1. Introduction

In this paper we consider the exploitation of gravity for tasks such as camera calibration and relative pose estimation for stereo systems. The basic idea is simple: if we take video sequences of objects being dropped or launched, then the image trajectories, combined with the assumption that the 3D trajectories follow the law of gravity, might provide us with information useful for estimating the imaging geometry. Two types of trajectory are of interest here: the trajectory of an object dropping down in a straight line (e.g. two or more such trajectories give us the vertical vanishing point) and that of an object being launched in a non vertical direction (i.e. with non zero horizontal velocity). In the latter case, the 3D trajectory is a parabola with vertical symmetry axis and we will see later that this information is useful to recover the vanishing geometry of stereo systems.

The paper is organized as follows. In §2, the camera model and some notations are introduced. The problem dealt with in this paper is described concretely in §3 and the geometric ideas leading to its solution are sketched in §4. In §§5 to 8, several ways of extracting information from image sequences of the described type are presented, which are useful for epipolar geometry estimation, calibration, pose estimation and also synchronization of cameras. Algorithms are briefly described in §9 and results of numerical simulations are presented in §10. Some practical issues are discussed in §11, followed by conclusions.

2. Background

Camera model. We use perspective projection to model cameras. A projection may be represented by a 3×4 matrix P that maps points of 3-space to points in 2-space: $\mathbf{q} \sim \mathsf{P}\mathbf{Q}$. Here, $\sim$ means equality up to a non zero scale factor, which accounts for the use of homogeneous coordinates. The projection matrix incorporates the so-called extrinsic and intrinsic camera parameters; it may be decomposed as:

$$\mathsf{P} \sim \mathsf{KR}\left(\; \mathtt{I}_3 \;\; | -\mathbf{t} \right) \;, \tag{1}$$

where $\mathtt{I}_3$ is the 3×3 identity matrix, R a 3×3 orthogonal matrix representing the camera's orientation, $\mathbf{t}$ a 3-vector representing its position, and K the 3×3 calibration matrix:

$$\mathsf{K} = \begin{pmatrix} \tau f & s & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \;. \tag{2}$$

In general, we distinguish 5 intrinsic parameters for the perspective projection model: the (effective) focal length f, the aspect ratio τ, the principal point (u_0, v_0) and the skew factor s accounting for non rectangular pixels.

Infinite homography. Consider the projections of a set of coplanar features in two images. The image features are linked by a 3×3 projective transformation, or homography. If the 3D features considered are located on the plane at infinity, the associated homography between the images is often referred to as the *infinite homography* [2]. This homography depends only on the two cameras' intrinsic parameters and their relative rotation, as follows:

$$\mathsf{H}_\infty \sim \mathsf{K}_2 \mathsf{R} \mathsf{K}_1^{-1} \tag{3}$$

If the infinite homography between two images can be determined, it provides constraints on the cameras' calibration and pose. These constraints have been used for example for calibrating a camera undergoing pure rotational displacements [1, 5, 10] or in stratified self-calibration approaches [4, 8]. In this paper, we will estimate the infinite homography using corresponding vanishing points and lines, i.e. projections of points or lines at infinity, and then use it for calibration.

3. Problem Description

We consider one or several static video cameras and our aim is to calibrate these and estimate their relative pose. We examine the potential use of gravity for these tasks. Input are video sequences of objects being dropped or launched. The videos consist thus of "snapshots" of the objects, at different times during their trajectory. From each snapshot, an image point is determined that is assumed to be the projection of the object's center of mass. In practice, a spheric object should be used; its projections are ellipses and the ellipse centers are usually a sufficiently good approximation for the projection of the center of mass if the object is not very close to the camera. Hence, the features we will use for calibration etc. are image points.

We assume that the cameras' frame rates are constant and identical. The frame rates need not be known however, and it is not required that the cameras are synchronized. Based on these simple assumptions, we examine how to use the image points and the knowledge that the corresponding 3D points lie on trajectories dictated by gravity, for calibration and pose estimation. Gravity gives us basically two useful pieces of information: first, all trajectories contain the same point at infinity (the vertical direction) and second, objects travel at constant horizontal velocity while their vertical velocity varies according to the law $v(t) = v(t = 0) - gt$. Once the vertical direction is known, this law allows us to compute ratios of point coordinates to e.g. compute the point of zero velocity of a linear trajectory (see below).

4. Geometric Ideas

Two types of trajectories are considered: linear trajectories, obtained by letting some object drop, and parabolic trajectories which are obtained when objects are launched. These trajectories are projected to lines and conics in the image plane. We will use the trajectories as a whole, but also correspondences between "snapshots" of the object taken during the trajectories.

Let us first consider a single camera which takes an image sequence of dropping objects. The objects drop vertically of course which means that their linear trajectories all have the same point at infinity (or direction). Consequently, the appearant trajectories in the image are a set of concurrent lines, the incidence point being the vanishing point corresponding to the vertical direction. Hence, we can determine this vanishing point. However, we can not obtain calibration constraints from linear trajectories, even if several cameras observe the trajectories simultaneously.

Consider now a camera observing several objects moving on parabolic trajectories. The 3D trajectories are parabolas with vertical symmetry lines, i.e. all those 3D parabolas contain the vertical point at infinity, as the linear trajectories. Hence, the conics obtained by projecting the trajectories all contain the vertical vanishing point. Interestingly, the tangent of any such conic at the vertical vanishing point is nothing else than the vanishing line of the motion plane, i.e. the plane supporting the 3D parabola (illustrated in figure 1). Hence, if we consider more than one camera, we can obtain correspondences of one vanishing point and several vanishing lines (one for each parabola) across different images. This alone is not sufficient to compute the infinite homography, since all the vanishing lines considered are concurrent. However, together with the epipolar geometry, two correspondences of vanishing lines are already enough to compute the infinite homography, and obtain the associated calibration constraints (described below).

The epipolar geometry can be estimated using individual snapshots of the objects, which can be used as point correspondences across the images. The projections of a minimum number of 3D points on at least two planes are sufficient to compute the epipolar geometry, i.e. two parabolic or three linear trajectories in general position are enough.

5. Using Linear Trajectories

If the camera is calibrated, it is simple to determine the horizon (the vanishing line of the horizontal planes) which might be used for example to orient a camera such that its gaze direction is horizontal. If $\mathbf{q}$ is the vertical vanishing point, then the horizon line $\mathbf{l}$ is given by:

$$\mathbf{l} \sim \mathsf{K}^{-\mathsf{T}} \mathsf{K}^{-1} \mathbf{q}$$

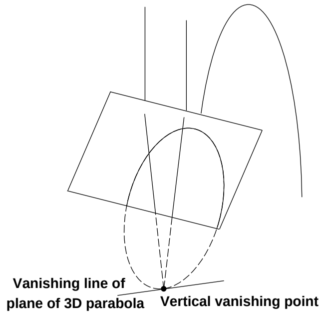

Figure 1. Example: a camera observing two linear and one parabolic trajectories. The image lines and conic all meet in the vertical vanishing point. The conic's tangent at this point is the vanishing line of its supporting plane. Dotted lines indicate primitives extending beyond the physical pixel grid, but lying on the mathematical image plane.

The camera could then be oriented by rotating it while tracking the horizon line, until it is horizontal and passes through the camera's principal point.

A second and maybe more interesting potential use of linear trajectories concerns the synchronization of cameras. Suppose we have determined the vertical vanishing points in the images, by intersecting the linear trajectories. It is then possible to determine, for each of the trajectories, the image point which corresponds to the position of zero velocity along the trajectory (three points on each trajectory are sufficient to do so; proof and formulas are omitted due to lack of space). In addition, we can determine the time of this event, given in frame numbers (naturally, the frame number will in general not be an integer as for the actually observed points). Hence, we can establish a correspondence of one (virtual) frame between the video sequences obtained from the cameras, which is enough to determine the time difference between them and thus to establish complete correspondence along the sequences.

This helps us to find point matches along the trajectories: given a point in one image, we can compute the frame number for another camera, which corresponds to the same time instant. The frame number will in general not be an integer, i.e. the corresponding image does not exist. However, using observed points on the trajectory, we can compute the position in the image that would have been observed at the required time instant (as above, three points and the frame numbers of the images where they were observed, are enough to do so). Hence, we are able to establish point correspondences between different cameras, even if they are not synchronized. These correspondences might be used to compute the epipolar geometry for example (points on at least three non-coplanar linear trajectories are needed).

6. Using Parabolic Trajectories

As described above, the projections of parabolic trajectories are conics which contain the vertical vanishing point. Important for us is that the conics' tangents at that vanishing point are the vanishing lines of the supporting planes of the 3D parabolas. Hence, observing parabolic trajectories in several cameras provides us with correspondences of vanishing lines. Given the epipolar geometry (which can be obtained as described in §5), two line correspondences are sufficient to compute the infinite homography. To see this, we note that the epipolar geometry allows us to establish point-wise correspondence among the vanishing lines, i.e. to produce valid correspondences of vanishing points. If we establish point correspondences on at least two vanishing lines, we already have enough constraints to compute the infinite homography. How to use the infinite homography for calibration and pose estimation is described in §§7 and 8.

We can thus obtain calibration constraints by observing parabolic trajectories in several views. What about single views: do the projections of parabolic trajectories in a single camera provide us with calibration constraints for that camera? The answer is no. The quintessence of the proof for this statement is that it is possible to obtain the affine structure of the parabolas (we know the vanishing line), but not more (any affine transformation that leaves the vertical direction fixed, maps a set of points on a parabolic trajectory to a set of points on another trajectory that also respects the law of gravity).

7. Calibration from the Infinite Homography

The infinite homography gives calibration constraints that have been used for calibrating cameras by rotating them about their optical centers [1, 5, 10] and for stratified self-calibration of cameras [4, 8]. There are a total of 5 calibration constraints: the 8 coefficients of the homography (9 minus 1 for the free scale factor) cover 3 parameters for the rotation (cf. equation (3)) and the remaining 5 constraints can be used to estimate the intrinsic parameters.

The constraints can be used in several ways. For example, it would be possible to "transfer" calibration information that is available for one camera, to the second camera: from equation (3), we obtain: $K_2K_2^T \sim H_\infty K_1^T K_1 H_\infty^T$ and from $K_2K_2^T$ we may obtain the calibration matrix K_2 using e.g. Choleski decomposition [9]. The most relevant practical situation however, is probably the case where the focal lengths of the two cameras have to be calibrated, given the other intrinsic parameters and assuming that the skew s is zero. We may separate the unknown from the known intrinsic parameters as:

$$K = \underbrace{\begin{pmatrix} \tau & 0 & u_0 \\ 0 & 1 & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{A} \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and then obtain from equations (3) and (2):

$$\left(A_2^{-1} H_\infty A_1\right) \begin{pmatrix} f_1^2 & 0 & 0 \\ 0 & f_1^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left(A_1^T H_\infty^T A_2^{-T}\right) \sim \begin{pmatrix} f_2^2 & 0 & 0 \\ 0 & f_2^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where all terms beside the two focal lengths are known. We determine the focal lengths by computing the symmetric matrix B that solves the following constraints to least squares:

$$\begin{aligned} B_{12} = B_{13} = B_{23} &= 0 \\ \left(XBX^T\right)_{12} = \left(XBX^T\right)_{13} = \left(XBX^T\right)_{23} &= 0 \\ B_{11} - B_{22} = \left(XBX^T\right)_{11} - \left(XBX^T\right)_{22} &= 0 \end{aligned}$$

where $X = \left(A_2^{-1} H_\infty A_1\right)$. From B and XBX^T we then obtain the two focal lengths in closed-form as $f_1^2 = 2\frac{B_{33}}{B_{11}+B_{22}}$ (and similarly for f_2^2).

Singularities. As in many (self-) calibration scenarios, there exist generic singularities in the form of relative pose configurations for which there is no unique solution for calibration. For the focal length calibration described above, the only singularity occurs when the optical axes of the two cameras are parallel (proof omitted due to lack of space). In that case, there are infinitely many pairs of values for the two focal lengths that are mathematically valid solutions.

Discussion. It is known that the focal lengths can be estimated from the epipolar geometry, without knowing the infinite homography [3, 7]. However, this problem is subject to more singularities: whenever the two optical axes are coplanar (i.e. the two cameras are fixated) and in some other cases, the calibration problem has no unique solution [7]. In practice, static stereo systems used for e.g. surveillance, will nearly always be approximately fixated, which will cause numerical instability for the calibration.

Another advantage of being able to use the infinite homography is that more calibration constraints are available: from the epipolar geometry, a maximum of 2 constraints on the intrinsic parameters can be extracted, whereas the infinite homography provides a maximum of 5 constraints.

8. Pose Estimation

Once calibration has been determined, there are several possibilities for estimating the relative pose between cameras, from the infinite homography, calibration and the epipolar geometry. The rotational component can be estimated e.g. using the infinite homography and calibration (cf. equation (3)). Another possibility is to extract pose information from the so-called essential matrix, which represents the calibrated epipolar geometry (see e.g. [6]).

9. Implementation

We briefly describe several parts of the implementation of the ideas described. A first version computes the vertical vanishing point in an image as the intersection of the available linear trajectories (obtained by fitting lines to points extracted from the projections of a falling object). Then, conics are fitted to obtain the projections of trajectories known to be parabolic. The tangents of these conics are computed and used, together with the epipolar geometry (computed from point matches as described in §5), to compute the infinite homography and then to calibrate.

An alternative algorithm does the computations in a much more direct manner and gives superior results: let $\mathbf{q}_i$ be the image coordinates of points arising from the *projection* of a parabolic trajectory. Let $\mathbf{Q}_i$ be points on an *arbitrary* 3D parabolic trajectory. We may compute an homography H that maps the $\mathbf{Q}_i$ to the $\mathbf{q}_i$. Interestingly, even if the $\mathbf{Q}_i$ do not represent the "true" 3D trajectory (e.g. the point of zero vertical velocity might occur at a different time instant), the homography H still enables computation of the true vanishing line, i.e. the line obtained by mapping the line at infinity by H is indeed the projection of the line at infinity of the true 3D plane of motion. If several images are used, correspondences of vanishing lines are thus established, and in addition, the homographies H allow to establish correspondences between vanishing points, which makes it unnecessary to compute the epipolar geometry to determine the infinite homography.

10. Experimental Results

We tested our algorithms by extensive numerical simulation. Here, we present results for the second algorithm of the previous section and the camera configuration shown in figure 2. Two parabolic trajectories (i.e. the minimum data) were created at random and image points were obtained by "sampling" the trajectory at time instants separated according to a frame rate of 25 fps. The intrinsic parameters of the cameras were $f = 10mm, \tau = 1, u_0 = v_0 = 256$, for a 512×512 image plane. The image points were perturbed by zero mean Gaussian noise of a standard deviation between 0 and 1 pixels. For different noise levels and elevation angles β (see figure 2), 100 random experiments each were performed. Figure 3 shows the median relative errors on the estimated focal lengths. In more than 60-95% of the experiments (depending on the noise level), the errors were below 10%. These results are rather encouraging, especially under the consideration that only linear algorithms were used (computation of homographies and focal lengths). With more than 2 trajectories and an optimization stage, the results should become significantly better.

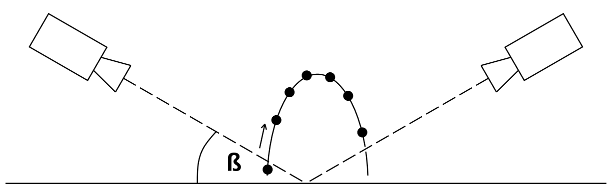

Figure 2. Configuration used for simulations.

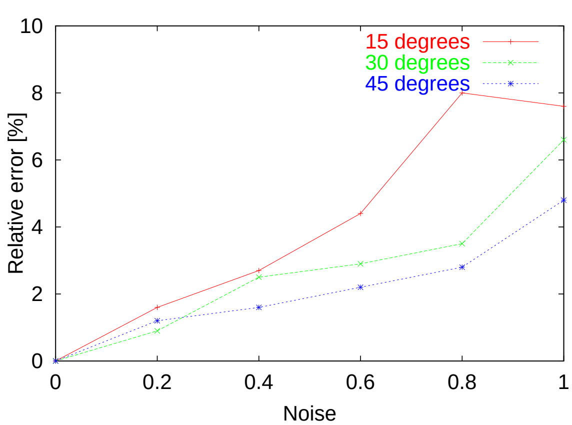

Figure 3. Errors of focal length calibration for different elevation angles and noise levels.

An elevation angle of 0° would mean that the cameras gaze at each other and, in particular, that the optical axes are parallel. This situation is singular (cf. §7), which was reflected by huge errors in the simulations (not shown in the graph). The graphs in figure 3 suggest that the results get better the further away the camera configuration is from the singularity, but there might be other reasons for this, e.g. for higher elevation angles, the vertical vanishing points are closer to the image centers, leading to a better conditioned computation of the infinite homography.

11. Practical Issues

To apply our approach in practice, we have to consider several issues. First of course, the frame rate of the cameras has to be high enough, i.e. the objects should be captured several times during their trajectories. The required frame rate depends mainly on the distance between the camera and the observed scene, and the camera's viewing angle. A simple computation shows that even in close-range conditions, a frame rate of 25 images per second is highly sufficient. Another issue is that we neglect air friction when making the assumption of 3D trajectories obeying perfectly the law of gravity. This should not introduce significant errors, since the most interesting trajectories will be those close to the vertex (point of zero vertical velocity), where the friction is minimal. Other critical issues, concerning the underlying image processing (extraction of objects in the images) are shutter speed and contrast. If the shutter speed is too low, motion blur is introduced, making the extraction of objects more difficult. Of course, sufficient contrast between the "calibration object" and the background is needed. This can be achieved by using fluorescent objects. The natural choice for calibration objects are spherical objects: they are perceived identically (if uniformly textured) independently of their orientation and the projections of their centers of gravity are well approximated as the centers of the ellipses forming the contours in the images.

12. Conclusions

We have presented ways of exploiting image sequences of moving objects if it is known that their trajectories obey the law of gravity. In particular, we have addressed the tasks of camera synchronization, computation of epipolar geometry, calibration and pose estimation. We mainly considered the calibration problem: while there are no useful constraints for single cameras, we obtain more constraints for stereo systems than would be possible without the exploitation of gravity. Maybe most importantly, the additional constraints suffer from fewer calibration singularities.

Especially, convergent (or fixated) two-camera systems do not represent a singular configuration, as it is for the case of focal length calibration from epipolar geometry alone.

This research is not mature enough yet for practical application. To prove its value, we start performing experiments with actual video sequences. The simulation results however, are encouraging and our approach might be useful e.g. for calibrating surveillance systems consisting of several cameras at distant locations.

Some parts of the discussion in this paper were given in informal style, due to lack of space. Proofs will be made available in a technical report.

References

[1] L. de Agapito, R.I. Hartley, E. Hayman, "Linear Self-Calibration of a Rotating and Zooming Camera," *IEEE Int. Conference on Computer Vision and Pattern Recognition*, pp. 15-21, 1999.

[2] O. Faugeras, "Stratification of Three-Dimensional Vision: Projective, Affine and Metric Representations," *Journal of the Optical Society of America A*, Vol. 12, pp. 465-484, 1995.

[3] R.I. Hartley, "Estimation of Relative Camera Positions for Uncalibrated Cameras," *European Conference on Computer Vision*, pp. 579-587, 1992.

[4] R.I. Hartley, "Euclidean Reconstruction from Uncalibrated Views," DARPA-ESPRIT *Workshop Applications of Invariants in Comp. Vis.*, pp. 187-202, 1993.

[5] R.I. Hartley, "Self-Calibration from Multiple Views with a Rotating Camera," *European Conference on Computer Vision*, pp. 471-478, 1994.

[6] H.C. Longuet-Higgins, "A Computer Program for Reconstructing a Scene from Two Projections," *Nature*, Vol. 293, pp. 133-135, 1981.

[7] G.N. Newsam, D.Q. Huynh, M.J. Brooks, H.P. Pan, "Recovering Unknown Focal Lengths in Self-Calibration: An Essentially Linear Algorithm and Degenerate Configurations," *XVIII* ISPRS-*Congress, Vienna, Austria*, Vol. XXXI, Part B3, 575-580, 1996.

[8] M. Pollefeys, L. van Gool, A. Oosterlinck, "The Modulus Constraint: A new Constraint for Self-Calibration," *International Conference on Pattern Recognition*, Vol. I, pp. 349-353, 1996.

[9] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C,* 2nd edition, Cambridge University Press, 1992.

[10] G.P. Stein, "Accurate Internal Camera Calibration Using Rotation, with Analysis of Sources of Error," *International Conference on Computer Vision*, pp. 230-236, 1995.