Multiple-Camera Tracking of Rigid Objects

Frédérick Martin, Radu Horaud

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Multiple-camera Tracking of Rigid Objects

Frédérick Martin — Radu Horaud

N° 4268

September 2001

THÈME 3

ISSN 0249-6399 ISRN INRIA/RR--4268--FR+ENG

INRIA
RHÔNE-ALPES

Multiple-camera Tracking of Rigid Objects

Frédérick Martin , Radu Horaud

Thème 3 — Interaction homme-machine,
images, données, connaissances

Projet Movi

Rapport de recherche n° 4268 — September 2001 — 30 pages

Abstract: In this paper we describe a method for tracking rigid objects using one or several cameras. The tracking process consists of aligning a 3-D model representation of an object with image contours by measuring and minimizing the image error between predicted model points and image contours. The tracker behaves like a visual servo loop where the internal and external camera parameters are updated at each new image acquisition. We study in detail the Jacobian matrix associated with this minimization process in the presence of both point-to-point and point-to-contour matches. We establish the minimal number of matches that are needed as well as the singular configurations leading to a rank-defficient Jacobian matrix. We find a mathematical link between the point-to-point and point-to-contour cases. Based on this link we show that the latter has the same kind of singularities than the former. Moreover, we study multiple camera configurations which optimize the robustness of the method in the presence of single-camera singularities, bad, noisy, or missing data. Extensive experiments done with a complex ship part and with up to three cameras validate the method. In particular we show that the tracker may well be used as a camera calibration procedure.

Key-words: object tracking, visual servoing, camera calibration

Work supported by the European Commission, Esprit Reactive LTR project VIGOR – Visually guided robots using uncalibrated cameras, grant number 26247.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Suivi d'objets rigides avec plusieurs caméras

Résumé : Nous décrivons une méthode de poursuite d'objets rigides utilisant une ou plusieurs caméras. Le processus de poursuite consiste en l'alignement d'un modèle 3-D représentant l'objet avec des contours image en mesurant et en minimisant l'erreur entre des points prédits à partir du modèle et des contours image. Le "trackeur" agit comme une boucle d'asservissement visuel et les paramètres internes et externes des caméras sont mis à jour pour chaque nouvelle aquisition d'images. Nous étudions en détail la matrice jacobienne associée au processus de minimisation pour des appariement point-point et point-contour. Nous établissons le nombre minimal d'appariements qui sont nécessaires de même que les configurations singulières conduisant à une jacobienne qui n'est pas de rang plein. Nous trouvons un lien mathématique entre les cas point-point et point-contour. Sur la base de ce lien nous montrons que les deux cas ont le même type de singularités. Nous étudions des configurations comportant plusieurs caméras qui optimisent la robustesse de la méthode en présence de singularités dues à une seule caméra, bruit, données manquantes ou données erronnées. Des expérimentations effectuées avec une pièce de bateau et avec trois caméras valident la méthode. En particulier, elle peut être utilisée comme une technique de calibration.

Mots-clés : suivi d'objet, asservissement visuel, calibration de caméra

Contents

1 Introduction, background, and approach

The problem of localizing and tracking moving objects using one or several cameras has been an active research topic. Objects fall into several categories, from rigid, deformable, articulated and rigid, to articulated and deformable. A common approach consists of using an object model and of estimating the position and orientation of this model such that some image-based error is minimized. The image error describes the discrepancy between measured object features and predicted model features. The ability to properly move the model such that it optimally corresponds to the actual object depends of a number of factors. Roughly speaking, the 3-D parameters associated with the object's behavior (motion parameters, shape deformation parameters, joint parameters for an articulated object, etc.) are related to the image error by a Jacobian matrix.

With only one camera there are inherent ambiguities and singularities. Indeed, one image configuration may lead to a number of 3-D actions and Jacobian singularities may lead to no action at all. Therefore, it may be advantageous to use several cameras instead of one. From a practical point of view this raises the problem of dealing with multiple camera geometry. It is more tedious to calibrate a multiple-camera system and to update the calibration data than to calibrate a single camera. Moreover, the vast majority of previous approaches consider the multi-camera system as a stereo device requiring that each object feature is viewed in at least two images and that image-to-image feature matches are provided.

In this paper we propose a multi-camera based method for tracking rigid objects. A geometric model of the tracked object is provided in advance. The method consists of a "visual servoing" approach applied to the object model: At each iteration of the tracking process the image error between model and object features allows to update the parameters associated with the position and orientation of the model.

First, we investigate the case of one camera. We establish the link between point-to-point tracking and point-to-contour tracking. We analyse both the cases of fixed and varying camera internal parameters. We study the singularities of the Jacobian matrix and we reveal the cases where single-camera tracking cannot be properly performed.

Second, we investigate the case of several cameras. We show how the single-camera case can be used to calibrate the multiple-camera layout. We establish the tracking formulation which allows a rigid camera layout with possibly varying internal camera parameters. This formulation consists of running in parallel several single-camera trackers and does not require any image-to-image correspondence. We show how most of the single-camera singularities can be avoided by using two or more cameras.

Third, we describe an implementation using three cameras and a complex rigid object. Each camera observes a different object part and therefore the system is very robust with

respect to partial occlusions simultaneously occurring in several images and with respect to total occlusion of one or two cameras.

1.1 Previous work

The idea of using pose for tracking stems from the work of Lowe [14] who used line-segment matches and the Levenberg-Marquardt non-linear minimization method. Tracking with variable internal camera parameters has been introduced by Kinoshita and Deguchi [11] who perform visual servoing and camera calibration *simultaneously*. Espiau performed an in-depths analysis of the convergence of visual servoing in the presence of varying focal length [7].

Armstrong and Zisserman [1] proposed to predict a model contour in the image and to search around this prediction for image points that are likely to lie on a matching contour. They apply this technique to the case of straight lines and use a robust method to fit a line to the image points. Drummond and Cipolla [5, 3, 6] showed that it is possible to cast the rigid model tracking problem into a linear problem using the Lie algebra of the rigid motion - the kinematic screw. They suggest a series of papers dealing with a complex 3-D object (a ship part) and with articulated objects. The approach of Drummond and Cipolla is interesting because it relies only weakly on point or line matches. Instead their technique relies on contour-to-points matching. They show that their approach can be used for calibrating a camera. Other similar approaches to tracking using an image prediction and searching in a window around this prediction can be found in [19] and [21].

All these approaches to tracking consider only one camera and put emphasis on the data association problem. Multi-view tracking has been barely investigated. There are several ways of using several cameras. One way is to consider the multi-camera setup as a 3-D sensor, capture 3-D measurements, and perform the tracking directly in 3-D space, [20], [2]. Another way is to use the epipolar constraint into the tracking loop itself and Lamiroy et al. [13] combine the tracking equations with the epipolar geometry constraint. Finally, when the cameras are far apart, there are very few image-to-image correspondences and single camera trackers may be performed in parallel.

In this paper we derive an explicit algebraic expression for the Jacobian matrix associated with point-to-contour correspondences. Although in the past robust results were obtained with this type of data association, none thoroughly studied the algebraic structure of the Jacobian, the minimal data sets necessary to run a tracker, and the possibly singular configurations which lead to tracking failures. Based on this analysis we are able to show that a multiple-camera approach based on point-to-contour correspondences lead to a robust rigid-object tracker that can deal with large occlusions.

1.2 Paper organization

The remaining of this paper is organized as follows. Section 2 reviews the problem of tracking a rigid object using a perspective camera model. The relantionship between the pose parameters (to be estimated by the tracker) and errors associated with the internal camera parameters is thoroughly investigated. Rigid object tracking is treated as a minimization problem and explicit formulae are derived for two types of image-to-model matches: point-to-point and point-to-contour assignments. Section 3 analyses the singularities associated with point-to-point assignments while section 4 provides a similar analysis for point-to-contour assignments. In particular, singular cases are revealed especially when the internal camera parameters are allowed to vary. Multiple-camera tracking is studied in section 5 where it is shown that the singularities associated with one camera disappear when two or more cameras are being used. Finally section 6 describes extensive experiments that illustrate all the variations of the tracking method.

2 Problem formulation

Without loss of generality we consider a moving rigid object observed by one camera. Tracking consists of a model being moved such that its apparent position and orientation with respect to the camera corresponds to the real position and orientation of the object. Let $\boldsymbol{s} = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_k)$ be a set of model points projected onto the image and let $\boldsymbol{s}^\star$ be the set of corresponding image points. The relationship between the model velocity and its apparent image velocity is:

$$\dot{\boldsymbol{s}} = \mathbf{J}\boldsymbol{T}$$

where $\boldsymbol{T}$ is the kinematic screw associated with the 3D model motion. The error between the predicted model position and the actual object position may be measured from their image projections:

$$\boldsymbol{e} = \mathbf{C}\,(\boldsymbol{s} - \boldsymbol{s}^\star)$$

Where $\mathbf{C}$ is combination matrix allowing to consider more measurements than the number of degrees of freedom (six in this case – three for the rotational velocity vector and three for the translational velocity vector). A common choice which insures convergences is to set $\mathbf{C} = \mathbf{J}^\top$.

The objective is to move the model such that the image error decreases:

$$\lambda \mathbf{J}^\top (\boldsymbol{s} - \boldsymbol{s}^\star) = -\mathbf{J}^\top \dot{\boldsymbol{s}} = -\mathbf{J}^\top \mathbf{J}\boldsymbol{T}$$

which allows as solution:

$$\boldsymbol{T} = -\lambda \left(\mathbf{J}^\top \mathbf{J}\right)^{-1} \mathbf{J}^\top (\boldsymbol{s} - \boldsymbol{s}^\star) \tag{1}$$

This is the basic visual servoing equation which is solved in order to maintain $\boldsymbol{s} - \boldsymbol{s}^\star$ as small as possible. In order to apply this formulation to the tracking problem, one may approximate the velocity screw by:

$$\boldsymbol{T} = \frac{d\boldsymbol{x}}{dt} = \frac{\boldsymbol{x} - \boldsymbol{x}_0}{t - t_0}$$

Hence, the equation above allows the incremental update of the pose parameters $\boldsymbol{x}$:

$$\boldsymbol{x} = \boldsymbol{x}_0 - \lambda(t - t_0) \left(\mathbf{J}^\top \mathbf{J} \right)^{-1} \mathbf{J}^\top (\boldsymbol{s} - \boldsymbol{s}^\star) \tag{2}$$

2.1 Camera model

The model-based object tracking just described assumes a pinhole camera model, i.e., an object point projects onto an image point using the well known camera projection matrix:

$$\tilde{\boldsymbol{m}} = \mathbf{K} \left[\begin{array}{cc} \mathbf{R} & \boldsymbol{t} \end{array} \right] \tilde{M} \tag{3}$$

where $\tilde{\boldsymbol{m}}$ is a homogeneous 3-vector describing the projective coordinates of an image point, $\mathbf{K}$ is the matrix of internal camera parameters, $\mathbf{R}$ is a rotation matrix, $\boldsymbol{t}$ is a translation vector, and $\tilde{M}$ is a homogeneous 4-vector describing the coordinates of a model point.

The internal camera parameters are f (horizontal and vertical scale factor, or focal length) and u_0 and v_0 (image coordinates of optical center):

$$\mathbf{K} = \left[\begin{array}{ccc} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{array} \right]$$

2.2 The image Jacobian

The 2-vector $\boldsymbol{m}$ describes the pixel coordinates associated with the projection $\tilde{\boldsymbol{m}}$. By taking the time derivatives we obtain the classical relationship:

$$\frac{d\boldsymbol{m}}{dt} = \mathbf{J}_m \left(\begin{array}{c} v_x \\ v_y \\ v_z \\ w_x \\ w_y \\ w_z \\ \frac{df}{dt} \\ \frac{du_0}{dt} \\ \frac{dv_0}{dt} \end{array} \right) \tag{4}$$

with the kinematic screw $\boldsymbol{T}^\top = (v_x\ v_y\ v_z\ w_x\ w_y\ w_z)$ and with:

$$\mathbf{J}_m = f \begin{bmatrix} -\frac{1}{z} & 0 & \frac{x}{z^2} & \frac{xy}{z^2} & -\left(1+\frac{x^2}{z^2}\right) & \frac{y}{z} & \frac{x}{zf} & \frac{1}{f} & 0 \\ 0 & -\frac{1}{z} & \frac{y}{z^2} & \left(1+\frac{y^2}{z^2}\right) & -\frac{xy}{z^2} & \frac{-x}{z} & \frac{y}{zf} & 0 & \frac{1}{f} \end{bmatrix} \tag{5}$$

where $(x\ y\ z)$ are the Euclidean coordinates of point M expressed in camera frame.

2.3 Camera model errors

The camera parameters may be unknown, partially known, or badly known. One important feature of the tracking algorithm described below is that it can accommodate with rough estimates of these parameters which are updated during tracking.

We analyse the effect of badly known camera parameters onto the pose parameters. Let $\mathbf{K}$ be the exact camera parameters and let $\mathbf{R}$ and $\boldsymbol{t}$ be the pose parameters associated with $\mathbf{K}$ and with the point correspondences $\boldsymbol{m} \leftrightarrow \boldsymbol{M}$. Let $\hat{\mathbf{K}}$ be an estimation of the true camera parameters. With these estimated parameters one can associate estimated pose parameters $\hat{\mathbf{R}}$ and $\hat{\boldsymbol{t}}$. The projection equation holds in both these cases:

$$\tilde{\boldsymbol{m}} = \hat{\mathbf{K}} \left[\hat{\mathbf{R}}\ \hat{\boldsymbol{t}}\right] \tilde{\boldsymbol{M}} = \mathbf{K} \left[\ \mathbf{R} \quad \boldsymbol{t}\ \right] \tilde{\boldsymbol{M}}$$

We have:

$$\tilde{\boldsymbol{m}} = \mathbf{K} \left[\ \mathbf{K}^{-1}\hat{\mathbf{K}}\hat{\mathbf{R}} \quad \mathbf{K}^{-1}\hat{\mathbf{K}}\hat{\boldsymbol{t}}\ \right] \tilde{\boldsymbol{M}}$$

By writing $\hat{\mathbf{K}} = \mathbf{K} + d\mathbf{K}$ and by first-order Taylor expansion we obtain, [10]

$$\mathbf{K'}^{-1}\mathbf{K} = \mathbf{I} - \begin{bmatrix} \frac{df}{f} & 0 & \frac{du_0}{f} \\ 0 & \frac{df}{f} & \frac{dv_0}{f} \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{I} + \mathbf{K}_\varepsilon \tag{6}$$

The pose parameters are affected by these internal parameter errors as follows:

$$\begin{aligned} \hat{\mathbf{R}} &= \mathbf{R} - \mathbf{K}_\varepsilon \mathbf{R} \\ \hat{\boldsymbol{t}} &= \boldsymbol{t} - \mathbf{K}_\varepsilon \boldsymbol{t} \end{aligned}$$

The estimation of f, u_0 and v_0 is affected by the value of f. For small focal lengths the internal parameters have equal importance. For large focal lengths, the accuracy of f is intrinsically more crucial than the accuracy of the optical center. If the focal center lies at approximatively the image center, the value of f is, on an average, 4-5 times greater than the values of u_0 and v_0. Therefore, the ability to track in the presence of variable focal length, estimate and correct its value on line is an important feature.

2.4 Tracking as a minimization problem

We consider now a more general tracking problem where the object is observed simultaneously by several cameras with possibly varying internal camera parameters. Let $\boldsymbol{q}$ denote the state-vector associated with such a setup: this vector encapsulates the pose parameters of the object with respect to a global camera centered frame as well as the internal parameters associated with the camera models. Therefore, the model predictions $\boldsymbol{s}$ are a function of $\boldsymbol{q}$.

One must distinguish between two cases:

1. Tracking based on point-to-point correspondences
2. Tracking based on point-to-contour correspondences

These two cases are illustrated on Figure 1. In the first case (point-to-point) the function to be minimized may be written as the sum of the Euclidean distances from the predicted model point $\boldsymbol{m}_i$ to its image match $\boldsymbol{m}_i^\star$:

$$Q_1(\boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{n} \|\boldsymbol{m}_i(\boldsymbol{q}) - \boldsymbol{m}_i^\star\|^2 = \frac{1}{2}\sum_{i=1}^{n} e_i^2$$

Therefore, the general form of the minimization problem is, in this case:

$$Q_1(\boldsymbol{q}) = \frac{1}{2}\boldsymbol{R}(\boldsymbol{q})^\top \boldsymbol{R}(\boldsymbol{q}) \tag{7}$$

In practice it is well known that it is difficult to obtain point-to-point correspondences. A current approach is to project a model contour onto the image using the current pose and camera parameters, search in a direction normal to the predicted model contour, find a corresponding image point lying onto a contour, and measure the distance to this point [17], [4]. For a projected model point $\boldsymbol{m}_i$ and a normal direction $\boldsymbol{n}_i$, this distance is approximated by, e.g., Figure 1:

$$d_i(\boldsymbol{q}) = \boldsymbol{n}_i^\top(\boldsymbol{q})(\boldsymbol{m}_i(\boldsymbol{q}) - \boldsymbol{m}_i^\star)$$

Notice that this formula is exact for straight-line contours and hold only approximatively for curved contours. The function to be minimized may be written as:

$$Q_2(\boldsymbol{q}) = \frac{1}{2}\left(\mathbf{N}(\boldsymbol{q})\boldsymbol{R}(\boldsymbol{q})\right)^\top \mathbf{N}(\boldsymbol{q})\boldsymbol{R}(\boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{n} d_i^2 \tag{8}$$

Matrix $\mathbf{N}$ of size $n \times 2n$ contains the normals to the projected