FREIE UNIVERSITÄT BERLIN

A Note on Calibration of Video Cameras for Autonomous Vehicles with Optical Flow

Ernesto Tapia, Raul Rojas

B-13-02
Februar 2013

FACHBEREICH MATHEMATIK UND INFORMATIK
SERIE B · INFORMATIK

A Note on Calibration of Video Cameras for Autonomous Vehicles with Optical Flow

Ernesto Tapia and Raúl Rojas

***Abstract*— We show how to compute the extrinsic parameters of a video camera from the optical flow measured in consecutive video frames. We assume that the camera is mounted on a car or a robot which move forward and sideways on a flat area to acquire the images used for the calibration. The vanishing points of the optical flow lines provide enough information to compute the camera rotation matrix. We also show how the flow lines on the ground together with the vehicle's velocity provide enough information to compute the camera position.**

I. Introduction

Our objective is to determine the orientation and position of a video camera from video frames alone, given the internal camera parameters. We assume that the camera is mounted on a vehicle or a robot, pointing more or less forward. The canonical position of the camera is such that the image plane is orthogonal to the direction going from the back to the front of the vehicle and also orthogonal to the ground, see Fig. 1.

The optical flow obtained in consecutive video frames is basic information we use to estimate the external parameters of the camera. Our main assumption is that the vehicle is moving on a flat area, so that a subset of the segments defined by the optical flow are the projection of parallel line segments of the world coordinate frame. This set of parallel lines intersect in the image plane at a common point, called the vanishing point. If the vehicle moves sideways when capturing the video frames, then the set of vanishing points define the horizon, which is also known as the vanishing line.

We will show how to estimate the direction and position of the camera, if we know the slope of the horizon line and the coordinates of the vanishing point of lines parallel to the forward movement of the vehicle.

II. Basic Concepts and Results

We will first introduce some useful concepts and results of projective geometry [2] that will facilitate our analysis.

Finding the external orientation and position of the camera consists in estimating the rotation matrix $\mathtt{R}$ and the translation vector $\mathbf{t}$ of the transformation

$$\mathbf{x} = \mathtt{K}[\mathtt{R}|\mathbf{t}]\mathbf{W}, \tag{1}$$

which maps a world vector $\mathbf{W}$ to a image vector $\mathbf{x}$ in the image plane. see Fig. 1.

The *internal camera matrix* $\mathtt{K}$ represents the internal orientation of the camera and transforms the projected space

This work was partially supported by the German Ministry of Education and Research (BMBF)

The authors are with the Institut für Informatik, Freie Universität Berlin, Arnimallee 7, 14195 Berlin, Germany

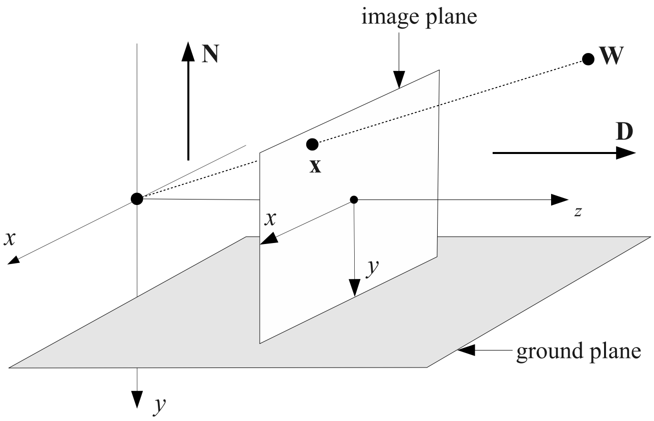

Fig. 1: Canonical camera position. The vehicle moves in the direction $\mathbf{D}$. The ground plane $y = h$ is orthogonal to the vector $\mathbf{N}$.

vectors into the camera coordinate frame. If $\mathtt{R} = \mathtt{I}$ and the $\mathbf{t} = \mathtt{0}$, we say that the camera is in its canonical position.

Actually, the transformation (1) is a linear mapping between homogeneous vectors. An *homogeneous vector* $\mathbf{X} = (X_1, \ldots, X_{n+1})^\top$ in the *projective space* $\mathbb{P}^n$ represents a vector $\mathbf{x} = (x_1, \ldots, x_n)^\top$ in $\mathbb{R}^n$, where $x_i = X_i / X_{n+1}$, $X_{n+1} \neq 0$ and $i = 1, \ldots, n$. The notation $(\mathbf{x}^\top, 1)^\top$ is homogeneous representation of the inhomogeneous vector $\mathbf{x}$. The world vector $\mathbf{W}$ belongs to $\mathbb{P}^3$ and the image vector $\mathbf{x}$ to $\mathbb{P}^2$.

A line $l_1 x + l_2 y + l_3 = 0$ in $\mathbb{R}^2$ is represented with the homogeneous vector $\mathbf{l} = (l_1, l_2, l_3)^\top$ in $\mathbb{P}^2$. Note that the point $\mathbf{x} = (x, y, 1)^\top$ belongs to the line, if and only if $\mathbf{x}^\top \mathbf{l} = 0$.

The *points at infinity* are represented in homogeneous coordinates as $(\mathbf{d}^\top, 0)^\top$. Geometrically, parallel lines and parallel planes in $\mathbb{P}^3$ intersect in points and lines formed with points at infinity. Thus, the point $\mathbf{D} = (\mathbf{d}^\top, 0)^\top$ represents the direction of parallel lines.

Proposition below is the main result we use to calculate the extrinsic parameters of the camera. This result describes how parallel lines and parallel planes in the world coordinate frame are transformed in the camera plane.

Proposition 1: The camera matrix $\mathtt{K}[\mathtt{R}|\mathbf{t}]$ maps the intersection of parallel planes with a common normal vector $\mathbf{N} = (\mathbf{n}^\top, 0)^\top$ to the vanishing line $\mathbf{l} = \mathtt{K}^{-\top}\mathtt{R}\mathbf{n}$. The camera matrix also transforms the intersection of parallel lines in the direction $\mathbf{D} = (\mathbf{d}^\top, 0)^\top$ to the vanishing point $\mathbf{v} = \mathtt{KR}\mathbf{d}$.

Next section shows how to compute the camera orientation if we know the horizon line defined by planes parallel to the ground plane and if we know the vanishing point of lines in the direction of the forward movement of the car. See Fig. 1.

III. External Camera Orientation and Position

Without lost of generality, we assume that camera coordinates are *normalized*. This assumption means that $\mathtt{K} = \mathtt{I}$ and $\mathbf{t} = \mathbf{0}$, so that (1) is simplified to a rotation

$$\mathbf{x} = \mathtt{R}\mathbf{w}, \tag{2}$$

where $\mathbf{w}$ is the inhomogeneous representation of the world vector $\mathbf{W} = (\mathbf{w}^\top, 1)^\top$. Note that we can always normalize the image vectors $\mathbf{x}$ by multiplying both sides of (1) with $\mathtt{K}^{-1}$.

A. Camera Orientation

We will represent the camera rotation $\mathtt{R}$ as the product

$$\mathtt{R} = \mathtt{R}_z(\gamma)\mathtt{R}_x(\alpha)\mathtt{R}_y(\beta), \tag{3}$$

where

$$\mathtt{R}_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_\alpha & -s_\alpha \\ 0 & s_\alpha & c_\alpha \end{pmatrix}, \tag{4}$$

$$\mathtt{R}_y(\beta) = \begin{pmatrix} c_\beta & 0 & s_\beta \\ 0 & 1 & 0 \\ -s_\beta & 0 & c_\beta \end{pmatrix}, \tag{5}$$

$$\mathtt{R}_z(\gamma) = \begin{pmatrix} c_\gamma & -s_\gamma & 0 \\ s_\gamma & c_\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

are the rotation matrices around the coordinate axis, and c_θ and s_θ are the cosine and sinus of the angle θ.

Suppose we know that the skyline is mapped to the line $\mathbf{l} = (l_1, l_2, l_3)^\top$ in the image plane. Given the vector $\mathbf{N} = (0, -1, 0, 0)^\top$ normal to the ground plane, Proposition 1 tells us that the line intersection of the planes orthogonal to $\mathbf{N}$ will be mapped to the line

$$\mathtt{R}(0, -1, 0)^\top = (c_\alpha s_\gamma, -c_\alpha c_\gamma, -s_\alpha). \tag{7}$$

Thus, the slope $\tan(\gamma)$ of the line (7) must be equal to the slope $-l_1/l_2$ of the line $\mathbf{l}$, because they represent homogeneously the same line. So we conclude

$$\gamma = \arctan(-l_1/l_2). \tag{8}$$

Now suppose we know that the set of lines parallel to $\mathbf{D} = (0, 0, 1, 0)^\top$are mapped to the vanishing point $\mathbf{v} = (v_1, v_2, 1)^\top$ in the image plane. Proposition 1 tell us that

$$\mathbf{v} \propto \mathtt{R}_z(\gamma)\mathtt{R}_x(\alpha)\mathtt{R}_y(\beta)(0, 0, 1)^\top, \tag{9}$$

so we conclude that the homogeneous vector

$$\mathtt{R}_x(\alpha)\mathtt{R}_y(\beta)(0, 0, 1)^\top = (-s_\beta, -c_\beta s_\alpha, c_\alpha c_\beta) \tag{10}$$

will be mapped to

$$\mathtt{R}_z^{-1}(\gamma)\mathbf{v} = (v_1 c_\gamma + v_2 s_\gamma, v_2 c_\gamma - v_1 s_\gamma, 1). \tag{11}$$

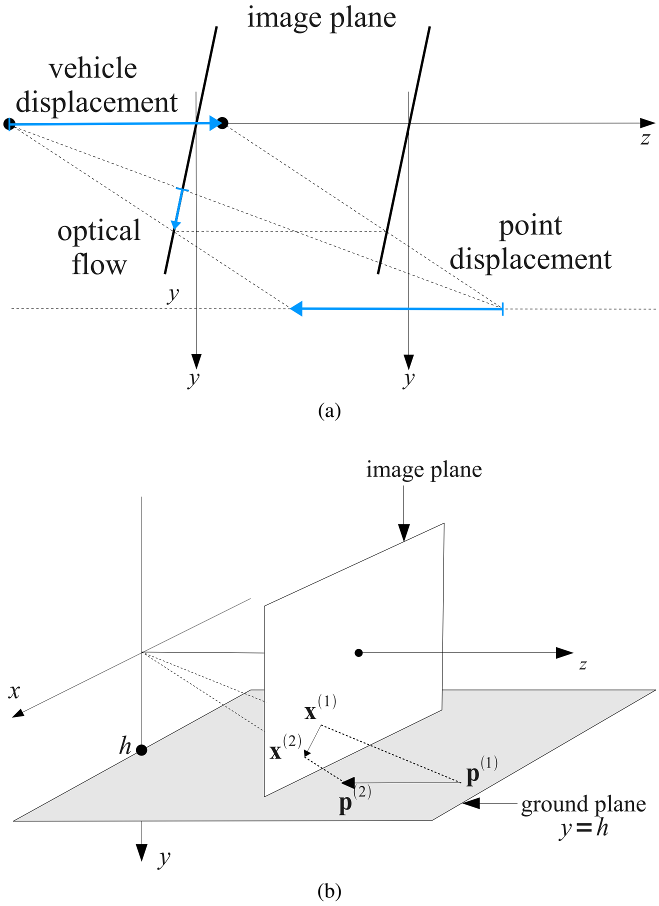

Fig. 2: (a) The translation of the vehicle on a flat area can be described by the optical flow between two consecutive video frames. (b) The ground plane $y = h$ is orthogonal to the vector $\mathbf{N}$. The points in the image plane $\mathbf{x}^{(i)}$ intersect the ground plane in the points $\mathbf{p}^{(i)}$.

Since the inhomogeneous form of (10) must be equal to (11), we have:

$$\alpha = \arctan(v_1 s_\gamma - v_2 c_\gamma) \tag{12}$$

$$\beta = \arctan(c_\alpha(v_1 c_\gamma + v_2 s_\gamma)). \tag{13}$$

Thus, we have found all the angles used by the rotation matrix.

B. Camera Position

We will show in this section how to compute the camera position relative to the ground. The main assumption here is that the points in the image plane are the mapping of points that belong to the ground, and that the images are captured during the forward motion of the vehicle. Under this assumption, the vehicle motion between two frames is equivalent to the motion of a world point on the ground and its motion on the image plane defined with its optical flow segment, see Fig. 2(a). Thus, if we assume that the ground plane is represented with the equation $y = h$, then the problem of computing the camera position relative to the ground reduces to computing the value h, see Fig. 2(b).

Remember that, in analogy to the two dimensional case, a plane $h_1 x + h_2 y + h_3 z + h_4 = 0$ in $\mathbb{R}^3$ is represented in homogeneous coordinates as $\mathbf{H} = (h_1, h_2, h_3, h_4)^\top$, and that a point $\mathbf{X} = (x, y, z, 1)$ belongs the plane if and only if $\mathbf{X}^\top \mathbf{H} = 0$.

The point $\mathbf{x}$ in the image plane can be *back projected* to the point $\mathbf{d} = \mathtt{R}^{-1}\mathbf{x}$ in the world space [2]. The line joining the origin with $\mathbf{d}$ will intersect the ground plane in the point $\mathbf{p} = \lambda \mathbf{d}$, so this point fulfills the equation

$$(\lambda \mathbf{d}^\top, 1)^\top (0, 1, 0, -h)^\top = 0. \tag{14}$$

By solving this equation for λ, one find the intersection point of the ray with the ground plane

$$\mathbf{p} = (\frac{hd_1}{d_2}, h, \frac{hd_3}{d_2}). \tag{15}$$

Now suppose that two points $\mathbf{x}^{(i)}$ in the image plane define a flow segment and that they are back projected to $\mathbf{d}^{(i)} = \mathtt{R}^{-1}\mathbf{x}^{(i)}$, $i = 1, 2$. We can use (15) to compute the displacement on the ground in terms of the image points:

$$\|\mathbf{p}^{(2)} - \mathbf{p}^{(1)}\| = |h| \sqrt{\left(\frac{d_1^{(2)}}{d_2^{(2)}} - \frac{d_1^{(1)}}{d_2^{(1)}}\right)^2 + \left(\frac{d_3^{(2)}}{d_2^{(2)}} - \frac{d_3^{(1)}}{d_2^{(1)}}\right)^2}, \tag{16}$$

where $\mathbf{p}^{(i)}$ are the intersection points of the rays defined by $\mathbf{d}^{(i)}$. This equation relates the horizontal displacement of the vehicle and the optical flow displacement, so if we know both displacements, then we can compute the height of the camera.

In practice we don't know the displacement on the ground, however, another quantity that is usually considered is the instantaneous velocity of the vehicle v. If we know the velocity v during the time interval Δt, then the displacement on the ground can be computed with $\|\mathbf{p}^{(2)} - \mathbf{p}^{(1)}\| = |v|\Delta t$ and the height of the camera with the formula

$$|h| = \frac{|v|\Delta t}{\sqrt{\left(\frac{d_1^{(2)}}{d_2^{(2)}} - \frac{d_1^{(1)}}{d_2^{(1)}}\right)^2 + \left(\frac{d_3^{(2)}}{d_2^{(2)}} - \frac{d_3^{(1)}}{d_2^{(1)}}\right)^2}}. \tag{17}$$

IV. Comments

The method described above also works if we have two straight parallel lines visible in the camera image which can be extrapolated up to their vanishing point. In this case we can orient the car parallel to the stright lines and determine the position of the vanishing point, without actual movement of the car. The direction of the parallel lines corresponds to the optical flow if the car was driving forward.

Finding the vanishing point of the optical flow can be useful for stereo vision; we can calibrate two cameras independently of each other and bring them later in correspondence. Note also that cameras pointing backwards can be calibrated in the same way as explained here, and even cameras where the vanishing point of the optical flow is not visible on the chip can be calibrated using the same approach. Given the position of the camera on the vehicle, the optical flow can be used as a poor mans IMU. This has been called "visual odometry", it can be more than just odometry.

It is also important to point out that optical flow algorithms are very general. They try to capture any kind of movement on the image. An optical flow algorithm specialized to determining the vanishing point could be faster and more accurate, since we only have to consider radial movement around the vanishing point. Radial distortion is still an issue, but at least we could limit the possible angles of the optical flow lines.

References

[1] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Inc., 2008.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[3] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, 2006.

[4] C. Tomasi and J. Shi. Good Features to Track. In *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 1994.