Determining the camera to robot-body transformation from mirror reflections

Anastasios I. Mourikis and Stergios I. Roumeliotis

December 7, 2007

1 Introduction

In this report, examine the case of a mobile robot moving in front of a planar mirror, while the robot’s camera observes the vehicle’s reflection on the mirror (cf. Fig. 1). The goal is to determine the 3D position of points on the robot’s body with respect to the camera, using the observations of the reflections of these points. We assume that the camera is calibrated both intrinsically, as well as extrinsically with respect to the robot’s odometry, and that the configuration of the mirror with respect to the robot is generally unknown.

2 Measurement model

Let \( \{C_j\} \) denote the camera frame at time-step \( j, j = 0, \ldots, N \). Since the camera-to-odometry transformation is known, we can use the odometry measurements to obtain estimates for the rotation and translation of the camera, \( c_o R_c^j \) and \( c_o p_c^j \), respectively, between frames \( \{C_o\} \) and \( \{C_j\} \).

Our goal is to compute the coordinates of \( M \geq 3 \) points on the robot body, expressed with respect to the camera frame, \( c_p^i, i = 1, \ldots, M \). Without loss of generality, we attach the world coordinate frame, \( \{W\} \), on the mirror, such that the mirror plane is identical to the global x-y plane. This frame assignment simplifies the expression of the coordinates of the reflection of any point on the mirror. Specifically, when all quantities are expressed in the world frame, the reflection of a point \( w p \) is a point \( w p' \), whose \( z \) coordinate is negated, i.e.,

\[
wp' = M wp, \quad \text{with} \quad M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}
\]  

At the \( j \)-th time instant, the vector from the origin of the camera frame to the reflection of \( c_p^i \), expressed in the camera frame, is:

\[
c_j^i = c_j^w R_w^w w p_i' + c_j^w p_w
\]

\[
= c_j^w R_w^w M (c_j^w R_w^w c_j^w p_i + w p_{c_j}) + c_j^w p_w
\]

where \( c_j^w R_w^w \) is the rotation matrix from the world frame to the camera frame, \( c_j^w p_w \) is the origin of the world frame expressed in the camera frame, and \( w p_{c_j} \) is the origin of the camera frame expressed in world coordinates. The latter is given by:

\[
w p_{c_j} = -c_j^w R_w^T c_j^w p_w
\]

Using this expression in (3), we obtain:

\[
c_j^i = (c_j^w R_w^w M c_j^w R_w^T) c_j^i - (c_j^w R_w^w M c_j^w R_w^T) c_j^w p_w + c_j^w p_w
\]

\[
= (c_j^w R_w^w M c_j^w R_w^T) c_j^i + (I - c_j^w R_w^w M c_j^w R_w^T) c_j^w p_w
\]
Figure 1: The robot moves in front of the mirror, and uses one of its cameras to observe a body-reference point \( c_p \) on its body.

where \( I \) denotes the \( 3 \times 3 \) identity matrix.

At this point, we express the unknown transformation between the camera frame \( \{ C_j \} \) and the world frame as a function of (i) the \textit{known} transformation between the frames \( \{ C_j \} \) and \( \{ C_o \} \), and (ii) the unknown transformation between \( \{ C_o \} \) and the world frame. The relations between these frames are described by:

\[
    c_j R_w = c_j R_{c_j} c_o R_w = c_o R_{c_j}^T c_o R_w = (c_j R_{c_j} c_o R_w) R_{c_j}^T
    \tag{7}
\]

\[
    c_j p_w = c_j p_{c_o} + c_o R_{c_j}^T c_o p_w
    \tag{8}
\]

Substitution in (6) yields:

\[
    c_j p_i = (c_j R_{c_j} c_o R_w M c_o R_{c_j}) c_o p_{c_o} + (I - c_o R_{c_j}^T c_o R_w M c_o R_{c_j}) (c_j p_{c_o} + c_o R_{c_j}^T c_o p_w)
    \tag{9}
\]

By defining the auxiliary matrix \( N = c_o R_w M c_o R_w^T \), and rearranging terms, the above expression is rewritten as:

\[
    c_j p_i = (c_j R_{c_j} N c_o R_{c_j}) c_o p_{c_o} + (I - c_o R_{c_j}^T N c_o R_{c_j}) c_o p_{c_o} + c_o R_{c_j}^T (I - N) c_o p_w
    \tag{10}
\]

We now turn our attention to the matrix \( N \). Note that

\[
    N = c_o R_w M c_o R_w^T
\]

\[
    = c_o R_w \left( I - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix} \right) c_o R_w^T
    \tag{11}
\]

\[
    = I - 2 c_o \hat{z}_w c_o \hat{z}_w^T
\]
Where $\hat{z}_w$ is the z-axis unit vector of the world frame, expressed with respect to the frame $\{C_o\}$. Due to our frame assignment, the z-axis of the world frame is perpendicular to the mirror plane. Thus, $\hat{z}_w$ is the normal vector to the mirror plane, expressed with respect to the first camera frame. For brevity, we employ the notation $\hat{z}_w = n$, and thus

$$N = I - 2nn^T \quad (12)$$

Substituting this expression into (10), we obtain:

$$c_j p_i = (I - 2c_o R^T_{c_j} n n^T c_o R_{c_j}) c_p_i + 2(c_o R^T_{c_j} n n^T c_o R_{c_j}) c_j p_{c_o} + 2c_o R^T_{c_j} n n^T c_o p_w$$

$$= c_o R^T_{c_j} ((I - 2nn^T) c_o R_{c_j} c_p_i + 2n^T c_o R_{c_j} c_j p_{c_o} + 2nn^T c_o p_w) \quad (13)$$

Using the identity $c_o R_{c_j} c_j p_{c_o} = c_o p_{c_j}$ and setting $d = n^T c_o p_w$, we obtain:

$$c_j p_i = c_o R^T_{c_j} ((I - 2nn^T) c_o R_{c_j} c_p_i - 2nn^T c_o p_{c_j} + 2d n) \quad (14)$$

In this expression, the quantities $c_o R^T_{c_j}$ and $c_o p_{c_j}$, which depend on the camera egomotion, are known. The unknown parameters are: (i) the feature position with respect to the camera, $c_p_i$, and (ii) the mirror-plane configuration, which is expressed by the unit vector $n$ and the scalar $d$ (note that $d$ is the distance of the origin of the frame $\{C_o\}$ from the mirror). These correspond to six unknown degrees of freedom: three for the feature position, two for the unit vector $n$, and one for $d$.

Every time the camera observes the reflection of the point $c_p_i$, we obtain a measurement of the unit vector along the direction of $c_j p_i$, i.e.:

$$c_j r_i = \alpha_{ij} c_j p_i \quad (15)$$

where $\alpha_{ij}$ is a normalizing constant to ensure unit length. Clearly, each such measurement provides two independent constraints. Hence, given at least three images of the same point from different locations, it is possible to determine both the coordinates of $c_p_i$, as well as the mirror configuration, $(n, d)$, with respect to the camera. This process can be further simplified by additionally tracking three fixed points, $m_i$, $i = 1; 2; 3$, on the mirror (e.g., the corners of the mirror, or markers placed on the mirror surface, cf. Fig. 1). When these points are detected in at least two images, we can employ triangulation to solve for their 3D positions and thus compute the parameters $n$ and $d$, which define the mirror configuration. In that case, (14) contains only three unknowns, corresponding to the point $c_p_i$ and therefore just two images suffice for computing its position estimate.