A Closed Form Solution for the Self-Calibration of Heterogeneous Sensors

Alessio Del Bue
alessio.delbue@iit.it
Marco Crocco
marco.crocco@iit.it
Igor Barros Barbosa
igorbb@gmail.com
Vittorio Murino
vittorio.murino@iit.it

Pattern Analysis & Computer Vision - PAVIS
Istituto Italiano di Tecnologia - IIT
Via Morego, 30, 16163 Genova, Italy

We present a novel closed-form solution for the joint self-calibration of video and range sensors solely from measurements as shown in Fig. 1. The approach single assumption is the availability of synchronous time of flight (i.e., range distances) measurements and visual position of the target on images acquired by a set of cameras. In such case, we make explicit a rank constraint that is valid for both image and range data. This rank property is used to find an initial and affine solution via bilinear factorization, which is then corrected by enforcing the metric constraints characteristic for both sensor modalities (i.e., camera and anchors constraints). The output of the algorithm is the identification of the target/range sensor 3D position and the calibration of the cameras.

Let us consider m range sensors and n point-like targets laying in a 3D space. Assuming no measurement errors, the following equations hold:

$$s_{i1}^2 + s_{i2}^2 + s_{i3}^2 + t_{j1}^2 + t_{j2}^2 + t_{j3}^2 - 2s_{i1}t_{j1} - 2s_{i2}t_{j2} - 2s_{i3}t_{j3} = d_{i,j}^2 \quad (1)$$

for $i = 1...m$, $j = 1...n$, where s_{il}, t_{jl} and $d_{i,j}$ denote respectively the sensor and target coordinates and the measured distance among them. By centering the sensors and target coordinates to the first sensor and the first target, the six quadratic terms in (1) disappear and a bilinear form can be obtained [1]:

$$-2\tilde{\mathtt{S}}\tilde{\mathtt{T}} = \tilde{\mathtt{D}}. \quad (2)$$

where $\tilde{\mathtt{S}}$, $\tilde{\mathtt{T}}$ and $\tilde{\mathtt{D}}$ matrices have dimension $(m-1)\times 3$, $3\times(n-1)$ and $(m-1)\times(n-1)$ respectively. Analogously, let us consider c affine cameras displaced in 3D space. Assuming an ideal projection of the n targets in the cameras frames, the following equations hold:

$$\mathbf{g}_{jk} = \begin{pmatrix} u_{jk} \\ v_{jk} \end{pmatrix} = \left[\begin{array}{c|c} \mathtt{R}_k & \mathbf{z}_k \end{array}\right] \begin{pmatrix} t_{j1} \\ t_{j2} \\ t_{j3} \\ 1 \end{pmatrix} = \mathtt{G}_k \begin{pmatrix} t_{j1} \\ t_{j2} \\ t_{j3} \\ 1 \end{pmatrix} \quad (3)$$

for $k = 1...c$, $j = 1...n$, where u_{kj} and v_{kj} represents the two image coordinates of the target j as seen by camera k. The 2×3 matrix $\mathtt{R}_k$ and the 2-vector $\mathbf{z}_k$ are the parameters of the cameras. By centering the target coordinates to the first target, Eq.(3) can be expressed in a matrix form as:

$$\tilde{\mathtt{G}} = \tilde{\mathtt{C}}\tilde{\mathtt{T}} \quad (4)$$

where matrices $\tilde{\mathtt{G}}$ and $\tilde{\mathtt{C}}$ have dimension $2c\times(n-1)$ and $2c\times 3$ respectively.

The common property for solving jointly the self-calibration problem is that both measured data sussist on a common subspace as defined by the target positions $\tilde{\mathtt{T}}$. The consequence is that the fusion of the modalities is for the first time strictly geometrical, in the sense that the data is now explicitly linked by the metric position of the targets. This leads to the possibility of computing a joint closed form solution using the range-visual constraints of the heterogeneous sensors. In particular Equations (2) and (4) can be merged together obtaining:

$$\mathtt{Y} = \begin{bmatrix} \tilde{\mathtt{D}} \\ \tilde{\mathtt{G}} \end{bmatrix} = \begin{bmatrix} -2\tilde{\mathtt{S}} \\ \tilde{\mathtt{C}} \end{bmatrix} \tilde{\mathtt{T}}. \quad (5)$$

The joint measurement matrix $\mathtt{Y}$ of size $(m+2c-1)\times(n-1)$ has rank equal to three since it is a product between a $(m+2c-1)\times 3$ matrix and a $3\times(n-1)$. If we apply a SVD to $\tilde{\mathtt{Y}}$ we have, in case of no noise, that the singular values after the third are equal to zero. Thus we can truncate these SVD components such as:

$$\mathtt{UVW} = \tilde{\mathtt{Y}}, \quad (6)$$

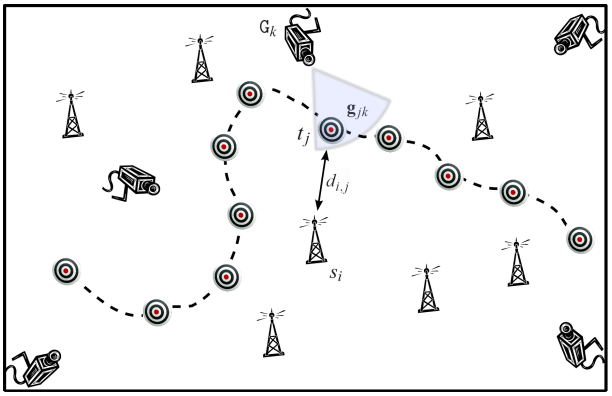

Figure 1: An example of the self-calibration problem for a heterogeneous sensor network. A target with 3D position $\mathbf{t}_j$ is measured by both a range sensor $\mathbf{s}_j$ and a video camera $\mathtt{G}_k$. Using just the scalar range distance d_{ij} from the sensors and the image coordinates of the target $\mathbf{g}_{jk}$ from the cameras, our algorithm recovers the 3D locations of the targets, sensors and it simultaneously calibrates each camera.

where $\mathtt{U}$ is an $(m-1)\times 3$ matrix, $\mathtt{V}$ is a 3×3 diagonal matrix and $\mathtt{W}$ is a $3\times(n-1)$ matrix. In a practical situation, in presence of measurement noise, the rank of $\tilde{\mathtt{Y}}$ will be higher than three: in this case only the three biggest singular values in $\mathtt{V}$ will be considered reducing the size of $\mathtt{U}$, $\mathtt{V}$ and $\mathtt{W}$ according to the noise-free case. From (5) and (6), for every invertible 3×3 matrix $\mathtt{C}$, the following relationships hold:

$$\begin{bmatrix} -2\tilde{\mathtt{S}} \\ \tilde{\mathtt{C}} \end{bmatrix} = \mathtt{UQ}_j \quad \text{and} \quad \tilde{\mathtt{T}} = \mathtt{Q}_j^{-1}\mathtt{VW}.$$

The matrix $\mathtt{Q}_j$ is called the "mixing matrix" since it mixes the components obtained from the SVD in order to obtain the correct solution given the original sensors localization problem. The matrix $\mathtt{Q}_j$ can be found exploiting the linear constraints given by the a priori known positions of a subset of range sensors, named anchors, as well as the quadratic constraints inherent to the affine camera model. We show in the paper that such constraints can be merged together, finding $\mathtt{Q}_j$ as a Cholesky decomposition of a matrix $\mathtt{H}$, whose entries are found through a linear least squares procedure.

The application extent of our approach is broad and versatile. In fact, with the same framework, we can deal with, but not restricted to, two very different applications. The first is aimed at calibrating cameras and microphones deployed in an unknown environment. The second uses a RGB-D device to reconstruct the 3D position of a set of keypoints using the camera and depth map images. Synthetic and real tests show the algorithm performance under different levels of noise and configurations of target locations, number of sensors and cameras. Though geometrical approaches for self calibration of range and video sensors are present in literature as two distinct problems, to the authors knowledge, for the first time we have presented a new geometrical constraint for the fusion of information acquired from video cameras and range sensors.

[1] M. Crocco, A. Del Bue, and V. Murino. A bilinear approach to the position self-calibration of multiple sensors. *IEEE Transactions on Signal Processing*, 60(2):660–673, February 2012.