Relative Orientation: A novel Approach[1)]

Gerald Schweighofer and Axel Pinz

Institute of Electrical Measurement and Measurement Signal Processing

Graz University of Technology, Schiesstattg.14B, A-8010 Graz, Austria

Abstract:

This paper presents a novel parameterization of the Relative Orientation with only three *parameters. Instead of using all structure and motion parameters and performing a full bundle-adjustment we show that, using the* Object Space Error *as the cost function, the optimal estimator depends only on the three parameters of the rotation of the relative orientation. So the number of parameters to be optimized can be reduced to three.*

1 Introduction

The process of estimating the Relative Orientation is typically divided into two parts. First a direct and also fast algorithm is used to obtain an estimate of the unknown parameters. This first estimate is usually based on a subset of measured points (eg. Estimating a direct solution of the essential matrix is possible with just 5 points [6]), or on a linearized solution of the original problem (eg. 8 or more points). In both cases this estimated solution is not an optimal one. So the second step in all structure-from-motion algorithms is an optimization process, which minimizes a cost function based on the motion and structure parameters [8]. This process is slow, because there are many parameters to be optimized.

There is a number of different cost functions which are used to optimize the relative orientation between two calibrated cameras. [10] gives an overview of three well-known ones. Two of them, *Distance Between Points and Epipolar Lines* and *Gradient-Weighted Epipolar Errors* are based on approximations of a geometric error and can be parameterized by the five parameters of the essential matrix. The third one *Distance Between Points and Reprojections* is based on a geometric error, but the number of parameters is $3n + 5$, the parameters of the structure (3 coordinates per point) and the five parameters of the essential matrix. For the case of two cameras (eg. a moving calibrated stereo-rig), a first estimate is available from a minimum of three points. For each position of the rig the points are reconstructed (eg. [2]), and afterwards the absolute orientation for the two reconstructions is found ([3],[9]). Due to noise in the image the reconstructions are not perfect. So the solution to the absolute orientation will not be an optimal solution to the relative orientation problem.

1) This work was supported by the Austrian Science Foundation (FWF, project P15748).

A parameterization of the minimization problem with the six motion parameters for a stereo-rig is given in [9]. The minimum number of parameters for the relative orientation found in the literature is five (in the one camera case) or six (for the stereo setup).

We present in this paper a general parameterization of the relative orientation between two views in only three parameters. General means in this case, that the parameterization does not depend on the number of used cameras. We first give a problem formulation and notation for multi-camera-sensors in Section 2. Then we give a short review of the cost function based on a geometric error in Section 3. Afterwards we derive in Section 3.1 an optimal reconstruction, based on the used cost function. In Section 3.2 we show how to estimate the optimal translational part of the relative orientation from all the measurements in a direct way. Finally we give some experimental results for the proposed parameterization in Section 4.

2 Problem formulation / Notation

We assume m *perspective* cameras which are fixed on a rig. We assume further that we know the calibration of the cameras, which consists of the intrinsic parameters (K_j) and the extrinsic parameters (R_j, t_j) of each camera ($j = 1..m$). The extrinsic parameters specify the location and orientation of the camera relative to a common scene coordinate system $\mathcal{S}$.

A point X_i (in the world coordinate system) is measured in the images as

$$\mathbf{v}_{kji} \quad \propto R_{kj} X_i + t_{kj} + \epsilon_{kji} \tag{1}$$

Where the index $k = \{1, 2\}$ represents one of the two poses of the rig, the index $j = 1...m$ stands for one of m cameras, and the index $i = 1...n$ specifies the world points. (e.g. $\mathbf{v}_{219}$ is world point number 9 measured in the first camera of the rig, which is in the second pose). The added term ϵ_{kji} is the error which is made through the measurement process and is different for all measurements. The symbol $\propto$ denotes 'directly proportional', so $\mathbf{v}_{kji}$ are measured up to an unknown scale factor. The rotation and translation components of each camera can be estimated from the calibration and the pose of the rig ($\mathcal{R}^k$,t^k) as:

$$R_{kj} = R_j \mathcal{R}^k \quad t_{kj} = R_j t^k + t_j. \tag{2}$$

For one special case ($m = 3$, $k = 2$) the situation is visualized in fig. 1.
We discuss the *Relative Orientation* ($\mathcal{R}$,t) between two poses ($k = 1$ and $k = 2$) of the rig (with all its cameras) in some scene coordinate system $\mathcal{S}$. Without loss of generality we can set the first pose to be at the origin of $\mathcal{S}$:

$$\mathcal{R}^1 = I \quad t^1 = [0, 0, 0]^T \tag{3}$$

$$\mathcal{R}^2 = \mathcal{R} \quad t^2 = t. \tag{4}$$

Where $\mathcal{R}$ and t describe the relative orientation of the rig.
The task of estimating the *relative orientation* of a rig is "to find the motion parameters $\mathcal{R}$ and t and the structure parameters $\mathbf{X}_i$ which minimize a given *cost function*".

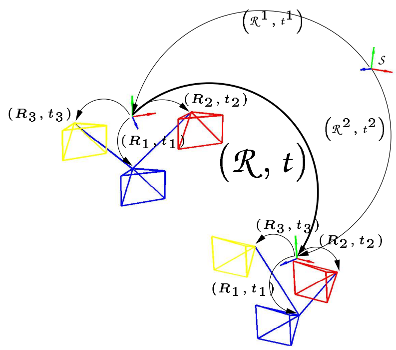


Figure 1: Relative Orientation for $m = 3$ and $k = 2$.

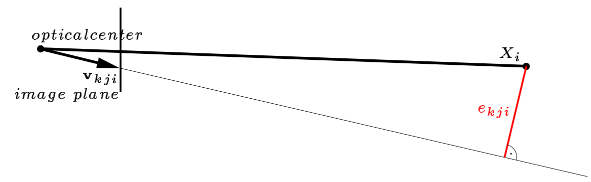


Figure 2: Scaled Object Space Error e_{kji}.

3 Cost function

In this Section we describe the cost function we used, which is the "Object Space Error"

$$e_{kji} = \|(I - V_{kji})(R_{kj}\mathbf{X}_i + \mathbf{t}_{kj})\|^2 \qquad \text{with} \qquad V_{kji} = \frac{\mathbf{v}_{kji}\mathbf{v}_{kji}^T}{\mathbf{v}_{kji}^T\mathbf{v}_{kji}}, \tag{5}$$

and was introduced by Lu et.al. [4]. As an example the error e_{kji} for one point X_i is shown in fig. 2. This error measures the distance between a world point X_i and the projection of this point onto the line of sight (which is defined by the optical center of the camera and and the measurement $\mathbf{v}_{kji}$). Based on this error we define now the solution to the *Relative Orientation Problem* as

$$\arg\min_{R,t,X_i} \sum_{k=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{m} e_{kji}(\mathcal{R}, t, X_i), \tag{6}$$

writing only the unknown parameters ($\mathcal{R}$,t,X_i) as parameters of the function (5). So the goal is to find those structure (X_i) and motion ($\mathcal{R}$ and t) parameters which minimize the total error. In the next section we derive an optimal reconstruction w.r.t. this problem and the chosen cost function (5).

3.1 Optimal structure reconstruction

Let us assume in this section that we know already the motion parameters $\mathcal{R}$ and t, which minimize eq. (6). We want to find the optimal reconstruction for the known motion parameters as

$$\arg\min_{X_i} \sum_{k=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{m} e_{kji}(\mathcal{R}, t, X_i) \rightarrow \textit{we know } \mathcal{R} \textit{ and } t \rightarrow \sum_{i=0}^{n} \arg\min_{X_i} \sum_{k=1}^{2}\sum_{j=l}^{m} e_{kji}(X_i). \tag{7}$$

So solving for an optimal X_i is given by $\quad \dfrac{\partial\left(\sum_{k=1}^{2}\sum_{j=1}^{m} e_{kji}(\mathbf{X}_i)\right)}{\partial \mathbf{X}_i} = \sum_{k=1}^{2}\sum_{j=1}^{m}\dfrac{\partial e_{kji}(\mathbf{X}_i)}{\partial \mathbf{X}_i} = 0. \quad (8)$

Writing the cost function (5) as a dot product

$$\begin{aligned} e_{kji} &= \left[(I - V_{kji})(R_{kj}\mathbf{X}_i + \mathbf{t}_{kj})\right]^T \left[(I - V_{kji})(R_{kj}\mathbf{X}_i + \mathbf{t}_{kj})\right] \\ &= \mathbf{X}_i^T R_{kj}^T VV_{kji} R_{kj}\mathbf{X}_i + \mathbf{t}_{kj}^T VV_{kji}\mathbf{t}_{kj} + 2\mathbf{t}_{kj}^T VV_{kji} R_{kj}\mathbf{X}_i, \end{aligned} \tag{9}$$

where $VV_{kji} = (I - V_{kji})^T(I - V_{kji})$, and differentiating e_{kji} (eg.(9)) with respect to $\mathbf{X}_i$ gives us

$$\frac{\partial e_{kji}}{\partial X_i} = 2(R_{kj}^T VV_{kji} R_{kj} \mathbf{X}_i + R_{kj}^T VV_{kji} \mathbf{t}_{kj}). \tag{10}$$

Using this result in eq. (8) gives us an equation for the optimal reconstruction of X_i:

$$\left(\sum_{k=1}^{2} \sum_{j=1}^{m} R_{kj}^T VV_{kji} R_{kj} \right) \mathbf{X}_i = - \sum_{k=1}^{2} \sum_{j=1}^{m} R_{kj}^T VV_{kji} \mathbf{t}_{kj}. \tag{11}$$

Eq. (11) is a linear system of equations, so $\mathbf{X}_i$ can be estimated in a direct way.

3.2 Optimal translation

We have seen in the previous section that we can estimate the optimal $\mathbf{X}_i$ w.r.t. the cost function (5) in a direct way. In this section we show that we can also estimate the optimal translation of the relative orientation t w.r.t. the cost function (5) in a direct way.
Let us assume here that we know already the rotational part of the relative orientation $\mathcal{R}$. Substituting the definition of $\mathbf{t}_{kj}$ (eq. (2) and eq. (3-4)) into eq. (11) gives us after some simple mathematics:

$$M_{A_i} \mathbf{X}_i = M_{B_i} t + \mathbf{v}_{C_i}. \tag{12}$$

We want to mention explicitly that the terms M_{A_i},M_{B_i} and $\mathbf{v}_{C_i}$, which are given in Appendix A, do not depend on the translation t. The solution $\mathbf{X}_i$ to eq. (12) can be written as

$$\mathbf{X}_i = \tilde{\mathbf{X}}_i t + \tilde{\mathbf{x}}_i \tag{13}$$

with $\tilde{\mathbf{X}}_i = M_{A_i}^{-1} M_{B_i}$ and $\tilde{\mathbf{x}}_i = M_{A_i}^{-1} \mathbf{v}_{C_i}$. This means that the optimally reconstructed world point $\mathbf{X}_i$, w.r.t. to the minimization problem eq. (6), can be estimated by the use of a projection $\tilde{\mathbf{X}}_i$ of t plus an offset $\tilde{\mathbf{x}}_i$. Eq. (13) gives us the possibility to write the optimally reconstructed point $\mathbf{X}_i$ as a function of t. Plugging this result into our minimization problem eq. (6) gives us

$$\arg\min_{\mathcal{R}, t} \sum_{k=1}^{2} \sum_{i=1}^{n} \sum_{j=1}^{m} e_{kji}(\mathcal{R}, t, X_i(\mathcal{R}, t)). \tag{14}$$

Using the vectorial form of the cost function (9) and the new representation of $\mathbf{X}_i$ (eq. (13)) we get

$$\begin{aligned} e_{kji}(\mathcal{R}, t) &= (\tilde{\mathbf{X}}_i t + \tilde{\mathbf{x}}_i)^T R_{kj}^T VV_{kji} R_{kj} (\tilde{\mathbf{X}}_i t + \tilde{\mathbf{x}}_i) + \mathbf{t}_{kj}^T VV_{kji} \mathbf{t}_{kj} \\ &+ 2\mathbf{t}_{kj}^T VV_{kji} R_{kj} (\tilde{\mathbf{X}}_i t + \tilde{\mathbf{x}}_i). \end{aligned} \tag{15}$$

Substituting the definition of $\mathbf{t}_{1j}$ and $\mathbf{t}_{2j}$ (eq. (2) and eq.(3-4)) gives us after some simplification the minimization problem

$$\arg\min_{\mathcal{R}, t} \left(t^T A(\mathcal{R}) t + 2\mathbf{b}(\mathcal{R})^T t + c(\mathcal{R}) \right), \tag{16}$$

which is the same as in eq. (14), but written as a polynomial in t. The "coefficients" $A(\mathcal{R})$, $\mathbf{b}(\mathcal{R})$ and $c(\mathcal{R})$ are given in Appendix B.

An optimal translation t_{opt} for this minimization problem is given by

$$\frac{\partial \text{ eq. (16)}}{\partial t} = 0 \Leftrightarrow \left(A(\mathcal{R})^T + A(\mathcal{R})\right) t = -\mathbf{b}(\mathcal{R}) \Leftrightarrow t_{opt} = -\left(A(\mathcal{R})^T + A(\mathcal{R})\right)^{-1} \mathbf{b}(\mathcal{R}) \quad (17)$$

Let us conclude:

To solve the *Relative Orientation Problem* we need to find the motion $\mathcal{R}$,t and all the structure parameters $\mathbf{X}_i$ to minimize the cost in eq. (6). We have shown above that we can estimate the optimal translation t_{opt} without the knowledge of the points $\mathbf{X}_i$ (we used only the $\tilde{\mathbf{X}}_i$ and $\tilde{\mathbf{x}}_i$ which only depends on $\mathcal{R}$). We have further shown that, given a known optimal t, we can estimate the optimal $\mathbf{X}_i$. So we can rewrite the *Relative Orientation Problem* as:

$$\arg\min_{\mathcal{R}} \sum_{k=1}^{2}\sum_{i=1}^{n}\sum_{j=1}^{m} e_{kji}(\mathcal{R}, t_{opt}(\mathcal{R}), X_{i_{opt}}(\mathcal{R}, t_{opt}(\mathcal{R})) \quad \equiv \quad \arg\min_{\mathcal{R}} fun_{ro}(\mathcal{R}) \quad (18)$$

with $$t_{opt}(\mathcal{R}) \xleftarrow{\text{is given by}} \text{eq. (20)} \quad \text{and} \quad X_{i_{opt}}(\mathcal{R}, t_{opt}(\mathcal{R})) \xleftarrow{\text{is given by}} \text{eq. (13)} \quad (19)$$

Equation (18) is the main result of this paper. It allows us to parameterize the cost function (6) as a nonlinear function fun_{ro}, which only depend on the rotation $\mathcal{R}$. Because any rotation is parameterizable with three parameters (eg. Euler angles or exponential parameterization [5]), the cost function itself is parameterizable with only three parameters.

Special Case $\mathbf{m} = \mathbf{1}$

In this case we have only *one* camera. Without loss of generality we assume that the coordinate system of the camera and the coordinate system of the rig coincide ($R_1 = I$ and $t_1 = [0,0,0]^T$). From eq. (23) we get that all $\mathbf{v}_{C_i} = [0,0,0]^T$. Based on this we see that all $\tilde{x}_i = [0,0,0]^T$. Further we get, by looking into eq. (25) and eq. (26), that $\mathbf{b}(\mathcal{R}) = [0,0,0]^T$ and $c(\mathcal{R}) = 0$. Finally the equation we need to solve to get an optimal solution for the translation t becomes in the case of one camera (eq. (16))

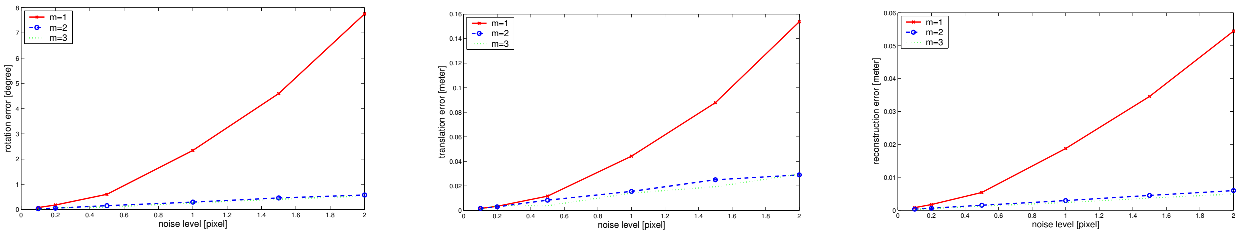
$$\arg\min_{\mathcal{R},t} \left(t^T A(\mathcal{R}) t\right). \quad (20)$$

A solution to this equation will be the trivial solution $t = [0,0,0]^T$, which minimizes the cost function and will not depend on $\mathcal{R}$. In the case of one camera and the definition of the cost function it is obvious that a translation vector $t = [0,0,0]^T$ will minimize the cost function. All structure parameters will become zero, they will coincide with the two camera centers. And so all lines of sights will go directly through them.

This happens because in this case (one camera) we can only obtain a reconstruction up to an unknown scale. To prevent us from getting the trivial solution we need to add the constraint $\|t\| = const$, eg. $\|t\| = 1$. So we need to solve the constrained system

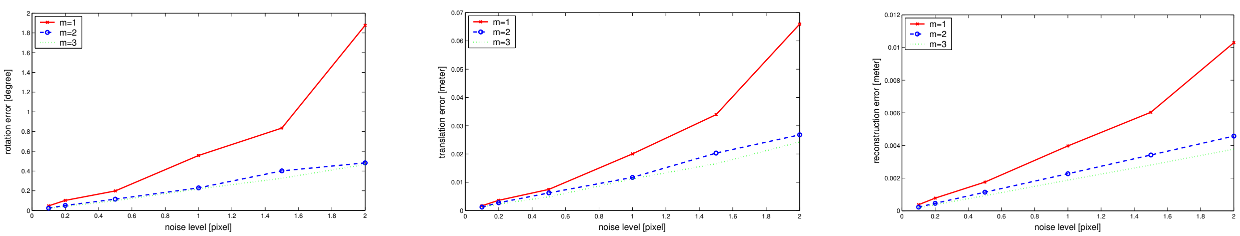
$$\arg\min_{\mathcal{R},t} \left(t^T A(\mathcal{R}) t\right) \quad with \quad \|t\| = 1. \quad (21)$$

Because we would like to minimize eq. (20) the solution for t is the *eigenvector* corresponding to the smallest *eigenvalue* of $\left(A(\mathcal{R})^T + A(\mathcal{R})\right)$.



(a) (b) (c)

Figure 3: Results of Experiment 1 (large, general displacement). Errors based on ground-truth: (a) Rotation error (b) translation error (c) reconstruction error for one, (solid), two (dashed) and three (dotted line) cameras.



(a) (b) (c)

Figure 4: Results of Experiment 2 (large, orbital displacement). Errors based on ground-truth: (a) Rotation error (b) translation error (c) reconstruction error for one (solid), two (dashed) and three (dotted line) cameras.

4 Experiments

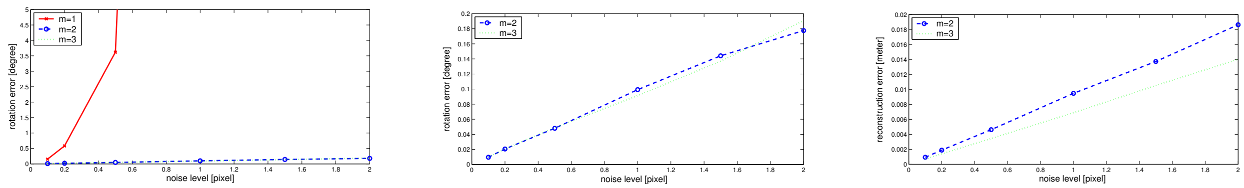
We tested the new parameterization on simulated data. The setup is as follows: Points are randomly distributed in a cube with one meter side length. The rig consists of 1 to max. 3 cameras. The camera centers are located at the points of an isosceles triangle with a side length of 25 cm. All three cameras look into the same direction (all three optical axes are parallel). The focal length of the cameras is set to 1000 pixels. All experiments are repeated 200 times for different noise levels, different numbers of cameras and random selection of points. For all experiments, the initial position of the rig is at $\mathcal{S} : \mathcal{R}^1 = I \; t^1 = [0, 0, 0]^T$, which is 3 meters from the center of the cube. The looking direction of the rig is also towards the center of the cube, i.e for the monocular case, the optical axis of the camera intersects with the center of the cube. We subsequently describe three different simulations for a large, general displacement, for a large displacement on an orbital trajectory, and for a pure in-plane translation.

Experiment 1 (*large, general displacement*): $\mathcal{R}^2 = R_x(35°)R_y(45°)$, $t^2 \approx [-2.12, 1.64, 2.01]^T$. This is a general case of relative orientation with a rather large translational displacement of approx. 3.35 meters and a significant rotation about two axes of the rig.

Experiment 2 (*large, orbital displacement*): This is a specific case of an orbital trajectory. The rig is rotated in the (x, z)-plane around the center of the cube for $90°$. Orbital motion sequences are known to be critical for monocular self-calibration and uncalibrated Euclidean reconstruction [7].

Experiment 3 (*small planar translation*): Another critical motion sequence ([7]) is planar motion (when optical axes and optical centers are moving an a single plane. We choose $\mathcal{R}^2 = I$, $t^2 = [-0.25, 0, 0]^T$.

To obtain comparable results we used 100 points for one ($m = 1$), 50 points for two ($m = 2$) and 33 points for three ($m = 3$) cameras. So the number of measured points is equal in all experiments.



(a) (b) (c)

Figure 5: Results of Experiment 3 (small planar translation). Errors based on ground-truth: (a) Rotation error. Zoomed error ($m = 2, 3$) for rotation (b) and reconstruction (c).

As initialization parameters for the optimization process we used for one camera ($m = 1$) the normalized eight-point-algorithm [1] and for more than one camera $m \geq 2$ the absolute orientation [3] between the two reconstructions.

Afterwards an optimization with our new parameterization is performed using a Quasi-Newton approach (we used the matlab command *fminunc*).

5 Results

Figure 3 presents the results for experiment 1 (large, general displacement). We show the result of the optimization process. Each point in the graph is the median over 200 experiments (the mean looks similar).

Figure 3(a) shows the error of the estimated rotation in degrees, fig. 3(b) shows the error of the estimated translation in meter. The structural error, which is the mean distance between reconstructed points (eg. (13) evaluated with the optimized rotation parameters), is shown in fig. 3(c). In general, the accuracy of the estimated motion and structure decreases with increasing noise level. Also the accuracy increases if we increase the number of used cameras. Furthermore, there is a rather linear relationship between error and noise for 2 and 3 cameras, while the error increases exponentially for the monocular case. However, even for one camera, we obtain less than 1° rotation error and less than 2 cm translation error for a noise level of 0.5 pixels.

Our experimental results for the two special cases 2 (fig. 4) and 3 (fig. 5) are in principle very similar: Linear relationship between motion and structure error and noise level for 2 or more cameras, with excellent reconstruction results. In the monocular case errors increase exponentially, with still quite good results for experiment 2, but experiment 3 constitutes a critical case.

6 Conclusion

We have presented a novel parameterization of the Relative Orientation with only *three* parameters. We have shown that, using the *Object Space Error* as the cost function, the optimal estimator depends only on the three parameters of the rotation of the relative orientation. This parameterization can be used for calibrated rigs with an arbitrary number of cameras. Experiments are given for one, two and three cameras, with excellent results for two or more cameras. Our new parameterization will be especially useful for applications of structure and motion analysis with a calibrated stereo rig.

References

[1] R.I. Hartley. In defence of the 8-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, June 1999.

[2] Richard Hartley and Peter Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[3] Berthold K.P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4:629–642, April 1987.

[4] C.P. Lu, G.D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):610–622, June 2000.

[5] Yi Ma, Stefano Soatto, Jana Košecká, and S. Shankar Sastry. *An Invitation to 3-D Vision*, chapter I.2.3. Springer-Verlag, 2004.

[6] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–777, June 2004.

[7] Peter Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pages 1100–1105, Washington, DC, USA, 1997. IEEE Computer Society.

[8] Bill Triggs, P. McLauchlan, Richard Hartley, and A. Fitzgibbon. Bundle adjustment – a modern synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag, 2000.

[9] Juyang Weng, Paul Cohen, and Nicolas Rebibo. Motion and structure estimation from stereo image sequences. *IEEE Transactions on Robotics and Automation*, 8(3):362–382, June 1992.

[10] Zhengyou Zhang. On the optimization criteria used in two-view motion analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(7):717–729, July 1998.

Appendix A: Coefficients for $\mathbf{X}_i$

$$M_{A_i} = \sum_{j=1}^{m} R_j^T VV_{1ji} R_j + \mathcal{R}^T R_j^T VV_{2ji} R_j \mathcal{R}\ ,\ M_{B_i} = -\sum_{j=1}^{m} \mathcal{R}^T R_j^T VV_{2ji} R_j \tag{22}$$

$$\mathbf{v}_{C_i} = \sum_{j=1}^{m} -\left(R_j^T VV_{1ji} + \mathcal{R}^T R_j^T VV_{2ji}\right) t_j \tag{23}$$

Appendix B: Coefficients for t

$$A(\mathcal{R}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \tilde{\mathbf{X}}_i^T R_j^T VV_{1ji} R_j \tilde{\mathbf{X}}_i + \tilde{\mathbf{X}}_i^T \mathcal{R}^T R_j^T VV_{2ji} R_j \mathcal{R} \tilde{\mathbf{X}}_i + R_j^T VV_{2ji}\left(2R_j \mathcal{R}\tilde{\mathbf{X}}_i + R_j\right) \tag{24}$$

$$\mathbf{b}(\mathcal{R})^T = \sum_{i=1}^{n}\sum_{j=1}^{m} \left(\tilde{\mathbf{x}}_i^T R_j^T + t_j^T\right) VV_{1ji} R_j \tilde{\mathbf{X}}_i + \left(\tilde{\mathbf{x}}_i^T \mathcal{R}^T R_j^T + t_j^T\right) VV_{2ji} R_j \left(\mathcal{R}\tilde{\mathbf{X}}_i + I\right) \tag{25}$$

$$c(\mathcal{R}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \left(\tilde{\mathbf{x}}_i^T R_j^T + 2t_j^T\right) VV_{1ji} R_j \tilde{\mathbf{x}}_i + \left(\tilde{\mathbf{x}}_i^T \mathcal{R}^T R_j^T + 2t_j^T\right) VV_{2ji} R_j \mathcal{R} \tilde{\mathbf{x}}_i + \sum_{k=1}^{2} t_j^T VV_{kji} t_j \tag{26}$$