ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-3, 2012
XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia

BUNDLE ADJUSTMENT FOR MULTI-CAMERA SYSTEMS WITH POINTS AT INFINITY

Johannes Schneider, Falko Schindler, Thomas Läbe, Wolfgang Förstner

University Bonn, Institute for Geodesy and Geoinformation
Department of Photogrammetry, Nussallee 15, 53115 Bonn, Germany
[johannes.schneider | falko.schindler]@uni-bonn.de, [laebe | wf]@ipb.uni-bonn.de, http://www.ipb.uni-bonn.de

KEY WORDS: bundle adjustment, omnidirectional cameras, multisensor camera, ideal points, orientation

ABSTRACT:

We present a novel approach for a rigorous bundle adjustment for omnidirectional and multi-view cameras, which enables an efficient maximum-likelihood estimation with image and scene points at infinity. Multi-camera systems are used to increase the resolution, to combine cameras with different spectral sensitivities (Z/I DMC, Vexcel Ultracam) or – like omnidirectional cameras – to augment the effective aperture angle (Blom Pictometry, Rollei Panoscan Mark III). Additionally multi-camera systems gain in importance for the acquisition of complex 3D structures. For stabilizing camera orientations – especially rotations – one should generally use points at the horizon over long periods of time within the bundle adjustment that classical bundle adjustment programs are not capable of. We use a minimal representation of homogeneous coordinates for image and scene points. Instead of eliminating the scale factor of the homogeneous vectors by Euclidean normalization, we normalize the homogeneous coordinates spherically. This way we can use images of omnidirectional cameras with single-view point like fisheye cameras and scene points, which are far away or at infinity. We demonstrate the feasibility and the potential of our approach on real data taken with a single camera, the stereo camera *FinePix Real 3D W3* from Fujifilm and the multi-camera system *Ladybug 3* from Point Grey.

1 INTRODUCTION

Motivation. The paper presents a novel approach for the bundle adjustment for omnidirectional and multi-view cameras, which enables the use of image and scene points at infinity, called "BACS" (Bundle Adjustment for Camera Systems). *Bundle adjustment* is the work horse for orienting cameras and determining 3D points. It has a number of favourable properties: It is statistically optimal in case all statistical tools are exploited, highly efficient in case sparse matrix operations are used, useful for test field free self calibration and can be parallelized to a high degree. *Multi-camera systems* are used to increase the resolution, to combine cameras with different spectral sensitivities (Z/I DMC, Vexcel Ultracam) or – like *omnidirectional cameras* – to augment the effective aperture angle (Blom Pictometry, Rollei Panoscan Mark III). Additionally multi-camera systems gain importance for the acquisition of complex 3D structures. *Far or even ideal points*, i. e. points at infinity, e. g. points at the horizon are effective in stabilizing the orientation of cameras, especially their rotation. In order to exploit the power of bundle adjustment, it therefore needs to be extended to handle multi-camera systems and image and scene points at infinity, see Fig. 1.

The idea. The classical collinearity equations for image points $\mathcal{X}'_{it}([x'_{it}; y'_{it}])$ of scene point $\mathcal{X}_i([X_i; Y_i; Z_i])$ in camera t with rotation matrix $\mathcal{R}_t([r_{kk'}])$ with k and $k' = 1, ..., 3$ and projection center $\mathcal{Z}_t([X_{0t}; Y_{0t}; Z_{0t}])$ read as

$$x'_{it} = \frac{r_{11}(X_i - X_{0t}) + r_{21}(Y_i - Y_{0t}) + r_{31}(Z_i - Z_{0t})}{r_{13}(X_i - X_{0t}) + r_{23}(Y_i - Y_{0t}) + r_{33}(Z_i - Z_{0t})} \quad (1)$$

$$y'_{it} = \frac{r_{12}(X_i - X_{0t}) + r_{22}(Y_i - Y_{0t}) + r_{32}(Z_i - Z_{0t})}{r_{13}(X_i - X_{0t}) + r_{23}(Y_i - Y_{0t}) + r_{33}(Z_i - Z_{0t})} \quad (2)$$

Obviously, these equations are not useful for far points or ideal points, as small angles between rays lead to numerical instabilities or singularities. They are not useful for bundles of rays of omnidirectional cameras, as rays perpendicular to the viewing direction, as they may occur with fisheye cameras, cannot be transformed into image coordinates. This would require different versions of the collinearity equation depending on the type of sensor as one would need to integrate the camera model into the bundle adjustment. Finally, the equations are not easily extensible to systems of multiple cameras, as one would need to integrate an additional motion, namely, the motion from the coordinate system of the camera system to the individual camera systems, which appears to make the equations explode.

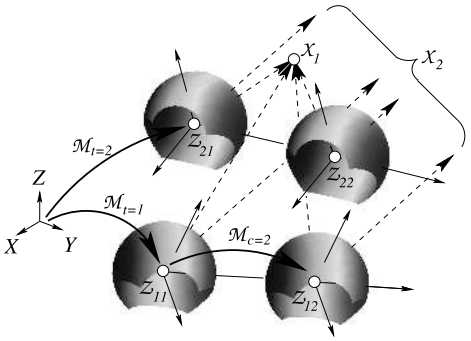


Figure 1: A two-camera system with Fisheye cameras $c = 1, 2$ with projection centers $\mathcal{Z}_{tc}$ and known motion $\mathcal{M}_c$ and unknown motion $\mathcal{M}_t$, having a field of view larger than 180° shown at two exposure times $t = 1, 2$ observing two points $\mathcal{X}_i, i = 1, 2$, one being close, the other at infinity. Already a block adjustment with a single camera moving over time will be stabilized by points at infinity.

Geometrically this could be solved by using homogeneous coordinates $\mathbf{x}'_{it}$ and $\mathbf{X}_i$ for image and scene points, a calibration matrix K_t and the motion matrix M_t in: $\mathbf{x}'_{it} = \lambda_{it}[\mathsf{K}_t \mid \mathbf{0}]\mathsf{M}_t^{-1}\mathbf{X}_i = \lambda_{it}\mathsf{P}_t\mathbf{X}_i$. Obviously, (a) homogeneous image coordinates allow for ideal image points, even directions opposite to the viewing direction, (b) homogeneous scene coordinates allow for far and ideal scene points, and including an additional motion is simply an additional factor.

However, this leads to two problems. As the covariance matrices $\mathbf{\Sigma}_{\mathbf{x}'_{it}\mathbf{x}'_{it}}$ of homogeneous vectors are singular, the optimization function of the Maximum Likelihood Estimation $\sum_{it} |\mathbf{x}_{it} -$

$\lambda_{it}\mathsf{P}_t\mathbf{X}_i|^2_{\Sigma^{-1}_{x'_{it}x'_{it}}}$ is not valid. A minor, but practical problem is the increase of the number of unknown parameters, namely the Lagrangian multipliers, which are necessary when fixing the length of the vectors $\mathbf{X}_i$. In large bundle adjustments with more than a million scene points this prohibitively increases the number of unknowns by a factor 5/3.

Task and challenges. The task is to model the projection process of a camera system as the basis for a bundle adjustment for a multi-view camera system, which consists of mutually fixed single view cameras, which allows the single cameras to be omnidirectional, requiring to explicitly model the camera rays and which allows for far or ideal scene points for stabilizing the configuration. The model formally reads as

$$x_{itc} = \mathcal{P}_{tc}(\mathcal{M}_t^{-1}(\mathcal{X}_i)) \tag{3}$$

with the I scene points $\mathcal{X}_i, i = 1, ..., I$, the T motions $\mathcal{M}_t, t = 1, ..., T$ of the camera systems from origin, the projection $\mathcal{P}_{tc}$ into the cameras $c=1, ..., C$ possibly varying over time, and the observed image points x_{itc} of scene point i in camera c at time/pose t.

For realizing this we need to be able to represent bundles of rays together with their uncertainty, using uncertain direction vectors, to represent scene points at infinity using homogeneous coordinates, and minimize the number of parameters to be estimated. The main challenge lies in the inclusion of the statistics into an adequate minimal representation.

Related Work. *Multi-camera systems* are proposed by many authors. E. g. Mostafa and Schwarz (2001) present an approach to integrate a multi-camera system with GPS and INS. Nister et al. (2004) discuss the advantage to use a stereo video rig in order to avoid the difficulty with the scale transfer. Savopol et al. (2000) report on a multi-camera system for an aerial platform for increasing the resolution. In all cases the multi view geometry is only used locally. *Orientation* of a stereo rig is discussed in Hartley and Zisserman (2000, p. 493). Mouragnon et al. (2009) proposed a bundle solution for stereo rigs working in terms of direction vectors, but they minimize the angular error without considering the covariance matrix of the observed rays. Frahm et al. (2004) present an approach for orienting a multi-camera system, however not applying a statistically rigorous approach. Muhle et al. (2011) discuss the ability to calibrate a multi-camera system in case the views of the individual cameras are not overlapping. Zomet et al. (2001) discuss the problem of re-calibrating a rig of cameras due to changes of the internal parameters. Bundle adjustment of camera systems have been extensively discussed in the thesis of Kim (2010). *Uncertain geometric reasoning* using projective entities has extensively been presented in Kanatani (1996), but only using Euclideanly normalized geometric entities and restricting estimation to single geometric entities. Heuel (2004) eliminating these deficiencies, proposed an estimation procedure which does not eliminate the redundancy of the representation and also cannot easily include elementary constraints between observations, see Meidow et al. (2009). The following developments are based on the minimal representation schemes proposed in Förstner (2012) which reviews previous work and generalizes e. g. Bartoli (2002).

2 CONCEPT

2.1 Model for sets of single cameras

2.1.1 Image coordinates as observations. Using homogeneous coordinates

$$\mathbf{x}'_{it} = \lambda_{it}\mathsf{P}_t\mathbf{X}_i = \lambda_{it}\mathsf{K}_t\boldsymbol{R}_t^{\mathsf{T}}[\boldsymbol{I}_3 \mid -\boldsymbol{Z}_t]\mathbf{X}_i \tag{4}$$

with a projection matrix

$$\mathsf{P}_t = [\mathsf{K}_t \mid \mathbf{0}_{3\times1}]\mathsf{M}_t^{-1}, \quad \mathsf{M}_t = \begin{bmatrix} \boldsymbol{R}_t & \boldsymbol{Z}_t \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5}$$

makes the motion of the camera explicit. It contains for each pose t: the projection center $\boldsymbol{Z}_t$ in scene coordinate system, i. e. translation of scene system to camera system, the rotation matrix $\boldsymbol{R}_t$ of scene system to camera system, and the calibration matrix K_t, containing parameters for the principal point, the principal distance, the affinity, and possibly lens distortion, see McGlone et al. (2004, eq. (3.149) ff.) and eq. (12). In case of an ideal camera with principal distance c thus $\mathsf{K}_t = \mathrm{Diag}([c, c, 1])$ and Euclidean normalization of the homogeneous image coordinates with the k-th row $\mathbf{A}_{t,k}^{\mathsf{T}}$ of the projection matrix P_t

$$\boxed{\mathbf{x}'^{e}_{it} = \frac{\mathsf{P}_t\mathbf{X}_i}{\mathbf{A}_{t,3}^{\mathsf{T}}\mathbf{X}_i} = \begin{bmatrix} \mathbf{A}_{t,1}^{\mathsf{T}}\mathbf{X}_i/\mathbf{A}_{t,3}^{\mathsf{T}}\mathbf{X}_i \\ \mathbf{A}_{t,2}^{\mathsf{T}}\mathbf{X}_i/\mathbf{A}_{t,3}^{\mathsf{T}}\mathbf{X}_i \\ 1 \end{bmatrix}} \tag{6}$$

we obtain eq. (1), e. g. $x'_{it} = \mathbf{A}_{t,1}^{\mathsf{T}}\mathbf{X}_i/\mathbf{A}_{t,3}^{\mathsf{T}}\mathbf{X}_i$.

Observe the transposition of the rotation matrix in eq. (4), which differs from Hartley and Zisserman (2000, eq. (6.7)), but makes the motion of the camera from the origin into the actual camera system explicit, see Kraus (1997).

2.1.2 Ray directions as observations. Using the directions from the cameras to the scene points we obtain the collinearity equations

$${}^k\mathbf{x}'_{it} = \lambda_{it}\,{}^k\mathsf{P}_t\mathbf{X}_i = \lambda_{it}\boldsymbol{R}_t^{\mathsf{T}}(\mathbf{X}_i - \mathbf{Z}_t) = \lambda_{it}[\boldsymbol{I}_3 \mid \mathbf{0}]\mathsf{M}_t^{-1}\mathbf{X}_i\,. \tag{7}$$

Instead of Euclidean normalization, we now perform spherical normalization $\mathbf{x}^s = \mathsf{N}(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$ yielding the collinearity equations for camera bundles

$$\boxed{{}^k\mathbf{x}'^{s}_{it} = \mathsf{N}({}^k\mathsf{P}_t\mathbf{X}_i)\,.} \tag{8}$$

We thus assume the camera bundles to be given as T sets $\{{}^k\mathbf{x}_{it}, i \in \mathcal{I}_t\}$ of normalized directions for each time t of exposure. The unknown parameters are the six parameters of the motion in ${}^k\mathsf{P}_t$ and the three parameters of each scene point. Care has to be taken with the sign: We assume the negative Z-coordinate of the camera system to be the viewing direction. The scene points then need to have non-negative homogeneous coordinate $X_{i,4}$, which in case they are derived from Euclidean coordinates via $\mathbf{X}_i = [\boldsymbol{X}_i; 1]$ always is fulfilled. In case of ideal points, we therefore need to distinguish the scene point $[\boldsymbol{X}_i; 0]$ and the scene point $[-\boldsymbol{X}_i; 0]$ which are points at infinity in opposite directions.

As a first result we observe: The difference between the classical collinearity equations and the collinearity equations for camera bundles is twofold. (1) The unknown scale factor is eliminated differently: Euclidean normalization leads to the classical form in eq. (6), spherical normalization leads to the bundle form in eq. (8). (2) The calibration is handled differently: In the classical form it is made explicit, here we assume the image data to be transformed into camera rays taking the calibration into account. This will make a difference in modelling the camera during self-calibration, a topic we will not discuss in this paper.

2.1.3 Handling far and ideal scene points. Handling far and ideal scene points can easily be realized by also using spherically

normalized coordinates for the scene points leading to

$$\boxed{{}^k\mathbf{x}_{it}'^s = \mathsf{N}({}^k\mathsf{P}_t\mathbf{X}_i^s)\,.} \tag{9}$$

Again care has to be taken with points at infinity.

2.2 Model for sets of camera systems

With an additional motion $\mathsf{M}_c(\boldsymbol{R}_c, \boldsymbol{Z}_c)$ for each camera of the camera system we obtain the general model for camera bundles

$$\boxed{{}^k\mathbf{x}_{itc}'^s = \mathsf{N}\left([\boldsymbol{I}_3 \mid \mathbf{0}_{3\times1}]\;\mathsf{M}_c^{-1}\;\mathsf{M}_t^{-1}\;\mathbf{X}_i^s\right)\,,} \tag{10}$$

making all elements explicit: the calibration of the camera system ($\mathsf{M}_c, c = 1,\dots,C$), assumed to be rigid over time, and the pose M_t of the system at time t. Substituting ${}^k\mathsf{P}_c = \boldsymbol{R}_c^\mathsf{T}[\boldsymbol{I}_3 \mid -\boldsymbol{Z}_c] = [\boldsymbol{I}_3 \mid \mathbf{0}_{3\times1}]\;\mathsf{M}_c^{-1}$ yields the model

$$\boxed{{}^k\mathbf{x}_{itc}'^s = \mathsf{N}\left({}^k\mathsf{P}_c\mathsf{M}_t^{-1}\mathbf{X}_i^s\right)} \tag{11}$$

with observed directions $\mathcal{X}_{itc}'({}^k\mathbf{x}_{itc}')$ represented by normalized 3-vectors, having two degrees of freedom, unknown or known scene point coordinates $\mathcal{X}_i(\mathbf{X}_i^s)$, represented by spherically normalized homogeneous 4-vectors, having 3 degrees of freedom and unknown pose $\mathcal{M}_t$ of camera system, thus 6 parameters per time. The projection matrices ${}^k\mathsf{P}_c$ only containing the relative pose of the cameras M_c are assumed to be given in the following.

2.3 Generating camera directions from observed image coordinates

In most cases the observations are made using digital cameras which physically are – approximately – planar. The transition to directions of camera rays need to be performed before starting the bundle adjustment. As mentioned before, this requires the internal camera geometry to be known. Moreover, in order to arrive at a statistically optimal solution, one needs to transfer the uncertainty of the observed image coordinates to the uncertainty of the camera rays. As an example we discuss two cases.

Perspective cameras. In case we have perspective cameras with small image distortions, we can use the camera-specific and maybe temporal varying calibration matrix

$$\mathsf{K}(\boldsymbol{x}', \boldsymbol{q}) = \begin{bmatrix} c & cs & x_H' + \Delta x(\boldsymbol{x}', \boldsymbol{q}) \\ 0 & c(1+m) & y_H' + \Delta y(\boldsymbol{x}', \boldsymbol{q}) \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

for the forward transformation

$${}^g\mathbf{x}' = \boldsymbol{K}(\boldsymbol{x}', \boldsymbol{q})\;{}^k\mathbf{x}'^s \tag{13}$$

from the observed image coordinates ${}^g\mathbf{x}'$, the ^{g} indicating the generality of the mapping system and the ray directions ${}^k\mathbf{x}'^s$. The calibration matrix besides the basic parameters, namely the principal distance c with image plane ${}^kZ = c$, shear s, scale difference m, and principal point $\boldsymbol{x}_H'$, contains additive corrections for modelling lens distortion or other deviations, which depend on additional parameters $\boldsymbol{q}$ and are spatially different via $\boldsymbol{x}$. In case of small deviations eq. (13) can easily be inverted. However, one must take into account, the different sign of the coordinate vector and the direction from the camera to the scene point, see Fig. 2,

$${}^k\mathbf{x}'^s \approx s\;\mathsf{N}\left(\mathsf{K}^{-1}({}^g\boldsymbol{x}', \boldsymbol{q})\;{}^g\mathbf{x}'\right) \tag{14}$$

with $s \in \{-1, +1\}$ such that ${}^kx_3'^s < 0$. This relation is independent of the sign of the third element of the calibration matrix.

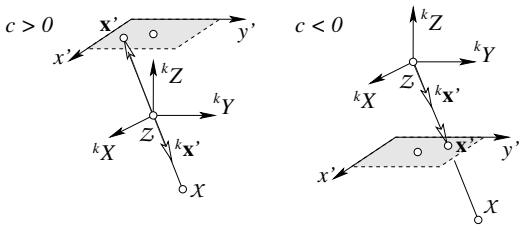


Figure 2: The direction of the homogeneous image coordinate vector and the direction of the ray is different depending on the sign of the principal distance c.

Given the covariance matrix $\Sigma_{x'x'}^g$ of the image coordinates, the covariance matrix of ${}^k\mathbf{x}'$ can be determined by variance propagation, omitting the dependency of the calibration matrix on the point coordinates $\boldsymbol{x}'$. Note that a point ${}^g\mathbf{x}'$ at infinity corresponds to the direction ${}^k\mathbf{x}'$ perpendicular to the viewing direction.

Omnidirectional single view point cameras. As an example for an omnidirectional single view camera we take a camera with a fisheye-lens. In the most simple case the normalized radial distance $r' = \sqrt{x'^2 + y'^2}/c$ of an image point with coordinates $[x', y']$ being the reduced image coordinates referring to the principal point, goes with the angle ϕ between the viewing direction and the camera ray. A classical model is the stereographic model

$$r' = c\,\tan\frac{\phi}{2}\,, \qquad \begin{bmatrix} {}^kx_1'^s \\ {}^kx_2'^s \\ {}^kx_3'^s \end{bmatrix} = \frac{1}{1 + r'^2/c^2}\begin{bmatrix} 2x'/c \\ 2y'/c \\ 1 - r'^2/c^2 \end{bmatrix}\,.$$

Again, the uncertainty of the image coordinates can be transformed to the uncertainty of the direction ${}^k\mathbf{x}'^s$ of the camera ray via variance propagation. In all cases the covariance matrix of the camera rays is singular, as the normalized 3-vector only depends on two observed image coordinates.

2.4 The estimation procedure

The collinearity equations in (11) contain three equations per observed camera ray and four parameters for the scene points, though, both being unit vectors. Therefore the corresponding covariance matrices are singular and more than the necessary parameters are contained in the equations. We therefore want to reduce the number of parameters to the necessary minimum. We do this after linearization.

Linearization and update for pose parameters. Linearization of the non-linear model leads to a linear substitute model which yields correction parameters which allow to derive corrected approximate values. We start with approximate values $\boldsymbol{R}^a$ for the rotation matrix, $\boldsymbol{Z}^a$ for the projection center, $\mathbf{X}^{sa}$ for the spherically normalized scene points, and $\mathbf{x}^a = \mathsf{N}(\mathsf{P}\;(\mathsf{M}^a)^{-1}\;\mathbf{X}^a)$ for the normalized directions. The Euclidean coordinates will be simply corrected by $\boldsymbol{Z} = \boldsymbol{Z}^a + \Delta\boldsymbol{Z}$, the three parameters $\Delta\boldsymbol{Z}$ are to be estimated. The rotation matrix will be corrected by premultiplication of a small rotation, thus by $\boldsymbol{R} = \boldsymbol{R}(\Delta\boldsymbol{R})\,\boldsymbol{R}^a \approx (\boldsymbol{I}_3 + \boldsymbol{S}(\Delta\boldsymbol{R}))\boldsymbol{R}^a$, where the small rotation $\boldsymbol{R}(\Delta\boldsymbol{R})$ depends on a small rotation vector $\Delta\boldsymbol{R}$ that is to be estimated.

Reduced coordinates and update of coordinates. The correction of the unit vectors is performed using reduced coordinates. These are coordinates, say the two-vector $\boldsymbol{x}_r$ of the direction $\mathbf{x}^s$, in the two-dimensional tangent space null $\left(\mathbf{x}^{sa\mathsf{T}}\right) = [\boldsymbol{r}, \boldsymbol{s}]$ of the unit sphere S^2 evaluated at the approximate values $\mathbf{x}^{sa}$

$$\boldsymbol{x}_r = \text{null}^\mathsf{T}\left(\mathbf{x}^{sa\mathsf{T}}\right)\,\mathbf{x}^s = \begin{bmatrix} \boldsymbol{r}^\mathsf{T}\mathbf{x}^s \\ \boldsymbol{s}^\mathsf{T}\mathbf{x}^s \end{bmatrix}\,. \tag{15}$$

The corrections $\Delta\boldsymbol{x}_r$ of these reduced coordinates are estimated. This leads to the following update rule

$$\mathbf{x}^s = \mathsf{N}\left(\mathbf{x}^{sa} + \text{null}\left(\mathbf{x}^{sa\mathsf{T}}\right)\,\Delta\boldsymbol{x}_r\right)\,. \tag{16}$$

Obviously, the approximate vector $\mathbf{x}^{sa}$ is corrected by

$$\Delta\mathbf{x} = \text{null}\left(\mathbf{x}^{sa\mathsf{T}}\right)\,\Delta\boldsymbol{x}_r \tag{17}$$

and then spherically normalized to achieve the updated values $\mathbf{x}^s$. Using (15) we now are able to reduce the number of equations per direction from three to two, making the degrees of freedom of the observed direction, being two, explicit. This results in pre-multiplication of all observation equations on (11) with $\text{null}^\mathsf{T}\left({}^k\mathbf{x}_{itc}^{sa\mathsf{T}}\right)$. When linearizing the scene coordinates we, following (17), use the substitution $\Delta\mathbf{X}_i^s = \text{null}\left(\mathbf{X}_i^{sa\mathsf{T}}\right)\,\Delta\boldsymbol{X}_{r,i}$. Then we obtain the linearized model

$$\begin{aligned} & {}^k\boldsymbol{x}_{r,itc} + \widehat{\boldsymbol{v}}_{x_{r,itc}} && (18)\\ = \;& -{}^k\boldsymbol{J}_s^\mathsf{T}\left({}^k\mathbf{x}_{itc}^a\right)\,\boldsymbol{R}_{tc}^{a\mathsf{T}}\,\boldsymbol{S}(\boldsymbol{X}_{i0}^a)\,\widehat{\Delta\boldsymbol{R}}_t \\ & +{}^k\boldsymbol{J}_s^\mathsf{T}\left({}^k\mathbf{x}_{itc}^a\right)\,\boldsymbol{R}_{tc}^{a\mathsf{T}}\,X_{ih}^a\widehat{\Delta\boldsymbol{Z}}_t && (19)\\ & +{}^k\boldsymbol{J}_s^\mathsf{T}\left({}^k\mathbf{x}_{itc}^a\right)\,{}^k\mathsf{P}_c\,(\mathsf{M}_t^a)^{-1}\,\text{null}\left(\mathbf{X}_i^{a\mathsf{T}}\right)\,\widehat{\Delta\boldsymbol{X}}_{ri} \end{aligned}$$

with

$${}^k\boldsymbol{J}_s({}^k\mathbf{x}_{itc}^a) = \frac{1}{|\mathbf{x}|}\left(\boldsymbol{I}_3 - \frac{\mathbf{x}\mathbf{x}^\mathsf{T}}{\mathbf{x}^\mathsf{T}\mathbf{x}}\right)\text{null}(\mathbf{x}^\mathsf{T})\Big|_{\mathbf{x}\,=\,{}^k\mathbf{x}_{itc}^a} \tag{20}$$

partitioning of the homogeneous vector $\mathbf{X}^s = [\boldsymbol{X}_0; X_h]$ and the compound rotation matrix $\boldsymbol{R}_{tc}^a = \boldsymbol{R}_t^a\boldsymbol{R}_c$ depending on $\widehat{\Delta\boldsymbol{R}}$, $\widehat{\Delta\boldsymbol{Z}}$, and $\widehat{\Delta\boldsymbol{X}}_r$.

We now arrive at a well defined optimization problem: find $\widehat{\Delta\boldsymbol{X}}_{r,i}, \widehat{\Delta\boldsymbol{R}}_t, \widehat{\Delta\boldsymbol{Z}}_t$ minimizing

$$\Omega\left(\widehat{\Delta\boldsymbol{X}}_{r,i}, \widehat{\Delta\boldsymbol{R}}_t, \widehat{\Delta\boldsymbol{Z}}_t\right) = \sum_{itc}\widehat{\boldsymbol{v}}_{r,itc}^\mathsf{T}\Sigma_{x_{r,itc}x_{r,itc}}^{-1}\widehat{\boldsymbol{v}}_{r,itc} \tag{21}$$

with the regular 2×2-covariance matrices

$$\Sigma_{x_{r,itc}x_{r,itc}} = {}^k\boldsymbol{J}_s^\mathsf{T}({}^k\mathbf{x}_{itc}^a)\begin{bmatrix}\Sigma_{x_{itc}x_{itc}} & \mathbf{0}\\ \mathbf{0}^\mathsf{T} & 0\end{bmatrix}{}^k\boldsymbol{J}_s({}^k\mathbf{x}_{itc}^a)\,. \tag{22}$$

3 EXPERIMENTS

3.1 Implementation details

We have implemented the bundle adjustment as a Gauss-Markov model in Matlab. Line preserving cameras with known inner calibration K_c as well as their orientation within the multi-camera system $\mathsf{M}_c(\boldsymbol{R}_c, \boldsymbol{Z}_c)$ are assumed to be known. To overcome the rank deficiency we define the gauge by introducing seven centroid constraints on the approximate values of the object points. This results in a free bundle adjustment, where the trace of the covariance matrix of the estimated scene points is minimal. Using multi-camera systems the scale is in fact defined by $\boldsymbol{Z}_c$. However the spatial extent of the whole block can be very large compared to this translation. We consider this by applying a weaker influence on the scale. We can robustify the cost function by down weighting measurements whose residual errors are too large by minimizing the robust Huber cost function, see Huber (1981). To solve for the unknown parameters, we can use the iterative Levenberg-Marquardt algorithm described in Lourakis and Argyros (2009).

Determination of approximate values. For initialization sufficiently accurate approximate values for object point coordinates and for translation and rotation of the multi-camera system at the different instances of time are needed. We use the results of the sift-feature based bundle adjustment *Aurelo* provided by Läbe and Förstner (2006) as approximate values for the pose of the multi-camera system. Object points are triangulated by using all corresponding image points that are consistent with the estimated relative orientations. Image points with residuals larger than 10 pel and object points behind image planes are discarded.

3.2 Test on correctness and feasibility

We first perform two tests to check the correctness of the implemented model and then show the feasibility on real data.

Parking lot simulation. We simulated a multi-camera system moving across a parking lot with loop closure, observing 50 scene points on the parking lot and 10 scene points far away at the horizon, i. e. at infinity (see Fig. 3). The multi-camera system contains three single-view cameras. Every scene point is observed by a camera ray on all 20 instances of time. The simulated set-up provides a high redundancy of observations. With a commonly

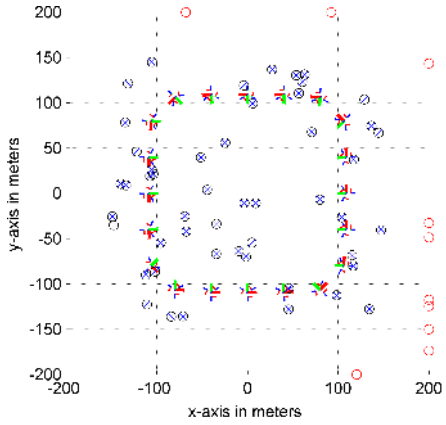


Figure 3: Simulation of a moving multi-camera system (poses shown as tripods) – e. g. on a parking lot – with loop closing. Scene points on the parking lot (crossed dots) and at the horizon (empty dots) being numerically at infinity are observed.

assumed standard deviation in the image plane of 0.3 pel and a camera constant with 500 pel, we add normally distributed noise with $\sigma_l = 0.3/500$ radiant on the spherically normalized camera rays to simulate the observation process. As initial values for the bundle adjustment we randomly disturb both the generated spherical normalized homogeneous scene points $\mathbf{X}_i^s$ by 10 % and the generated motion parameters $\boldsymbol{R}_t$ and $\boldsymbol{Z}_t$ of M_t by 3° and 2 meters respectively.

The iterative estimation procedure stops after six iterations, when the maximum normalized observation update is less than 10^{-6}. The residuals of the observed image rays in the tangent space of the adjusted camera rays, which are approximately the radiant of small angles, do not show any deviation from the normal distribution. The estimated a posteriori variance factor $\widehat{s}_0 = 1.0021$ approves the a priori stochastic model with variance factor $\sigma_0 = 1$. In order to test if the estimated orientation parameters and scene point coordinates represent the maximum-likelihood estimation for normally distributed noise on the observations, we have generated the same simulation 2000 times with different random noise. The mean of the estimated variance factors is not significantly different from one, indicating an unbiased estimator with minimum variance. These results confirm the correctness of the implementation.

Stereo camera data set. In order to test feasibility on real data we apply the bundle adjustment on 100 stereo images of a building with a highly textured facade taken with the consumer stereo camera *FinePix Real 3D W3* from Fujifilm (see Fig. 4). We use *Aurelo* without considering the known relative orientation between the stereo images to obtain an initial solution for the camera poses and the scene points. The dataset contains 284 813 image points and 12 439 observed scene points.

(a) left image (b) right image

Figure 4: Sample images of the stereo camera data set.

Starting from an a priori standard deviation for the image coordinates of $\sigma_l = 1$ pel, the a posteriori variance factor is estimated with $\widehat{s}_0 = 0.37$ indicating the automatically extracted Lowe points to have an average precision of approximately 0.4 pel. The resulting scene points and the poses of the right stereo camera images are shown in Fig. 5.

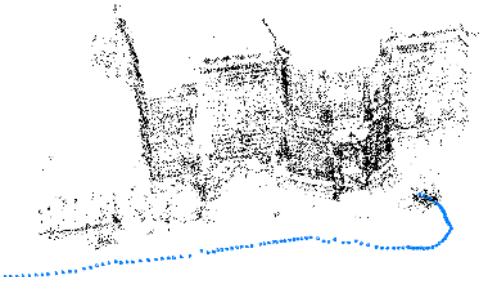


Figure 5: Illustration of the estimated scene points and poses of the right stereo camera images.

3.3 Decrease of rotational precision excluding far points

Bundle adjustment programs, such as *Aurelo*, cannot handle scene points with glancing intersections – e. g. with maximal intersection angles lower than $\gamma = 1$ gon – which therefore are excluded in the estimation process to avoid numerical difficulties. Far scene points, however, can be observed over long periods of time and therefore should improve the quality of rotation estimation significantly. We investigate the decrease of precision of the estimated rotation parameters of $\widehat{R}_t$ when excluding scene points with glancing intersection angles. In detail, we will determine the average empirical standard deviation $\sigma_{\alpha_t} = \widehat{s}_0 \sqrt{\operatorname{tr} \Sigma_{\widehat{R}_t \widehat{R}_t}/3}$ for all estimated rotation parameters and report the average decrease of precision by excluding far points determined by the geometric mean, namely $\exp\left[\sum_t^T \log(\sigma_{\alpha_t}/\sigma'_{\alpha_t})/T\right]$, where σ'_{α_t} represents the resulting average empirical standard deviation when scene points whose maximal intersection angle is lower than a threshold γ are excluded.

Parking lot simulation. We determine the decrease of precision for the estimated rotation parameters by excluding a varying number of scene points at infinity on the basis of the in Section 3.2 introduced simulation of a moving multi-camera system on a parking lot. Again we generate 50 scene points close to the multi-camera positions and vary the number of scene points at infinity to be 5, 10, 20, 50 and 100. The resulting average decrease in precision is 0.40 %, 2.31 %, 4.71 %, 7.03 % and 10.54 % respectively.

Multi-camera data set. We apply the bundle adjustment to an image sequence consisting of 360 images taken by four of the six cameras of the multi-camera system *Ladybug 3* (see Fig. 6). The *Ladybug 3* is mounted on a hand-guided platform and is triggered once per meter using an odometer. Approximate values are obtained with Aurelo by combining the individual cameras into a single virtual camera by adding distance dependent corrections on the camera rays, see Schmeing et al. (2011).

(a) c=1 (b) c=2 (c) c=3 (d) c=4

Figure 6: Sample images of the ladybug 3 data set.

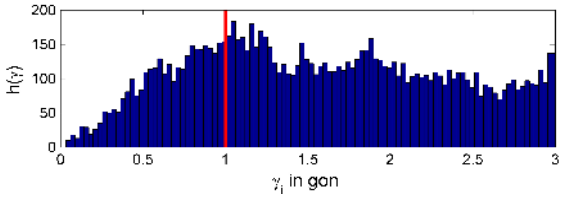
The data set contains 10 891 of 26 890 scene points observed with a maximal intersection angles per point significantly lower than $\gamma = 1$ gon, see the histogram in Fig. 7(a). The average standard deviation of each estimated rotation parameter is shown in Fig. 7(b) showing the individual gain in precision that is mainly obtained due to a higher number of observed scene points at the individual poses, as can be seen in the scatter plot in Fig. 7(c). Some of the cameras show very large differences in the precision, demonstrating the relevance of the far scene points in the *Ladybug 3* data set. Using far points results in an almost constant precision of the rotation parameters over all camera stations, in contrast to the results of the bundle adjustment excluding far points. The estimated a posteriori variance factor amounts $\widehat{s}_0 = 1.05$ using an a priori stochastic model with $\sigma_l = 1$ pel for the image points, indicating a quite poor precision of the point detection which mainly results from the limited image quality.

Urban drive data set. We make the same investigation on an image sequence consisting of 283 images taken by a single-view camera mounted on a car. The data set contains 33 274 of 62 401 scene points observed with a maximal intersection angle per point smaller than $\gamma = 1$ gon, see Fig. 8(a). Excluding those scene points decreases the average precision of the estimated rotation parameters by about 17.41 %. The average standard deviation of each estimated rotation parameter is shown in Fig. 8(b) showing the individual gain in precision that again is mainly obtained due to a higher number of observed scene points at the individual poses, shown in scatter plot of Fig. 8(c). The estimated posteriori variance factor amounts $\widehat{s}_0 = 0.54$ using a priori stochastic model with $\sigma_l = 1$ pel for the image points, indicating the precision to be in a normal range.

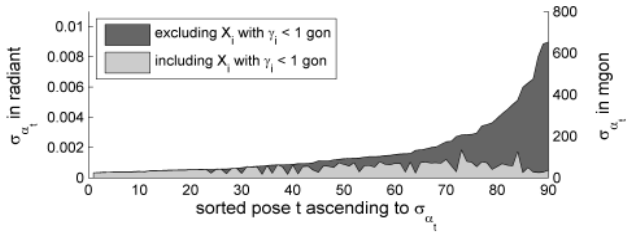
4 CONCLUSIONS AND FUTURE WORK

We proposed a rigorous bundle adjustment for omnidirectional and multi-view cameras which enables an efficient maximum-likelihood estimation with image and scene points at infinity. Our experiments on simulated and real data sets show that scene points at the horizon stabilize the orientation of the camera rotations significantly. Future work will focus on solving the issue of estimating scene points like a church spire which is near to some cameras and for some others lying numerically at infinity. Furthermore we are developing a fast C-implementation and a concept for determining approximate values for the bundle adjustment of multi-view cameras.

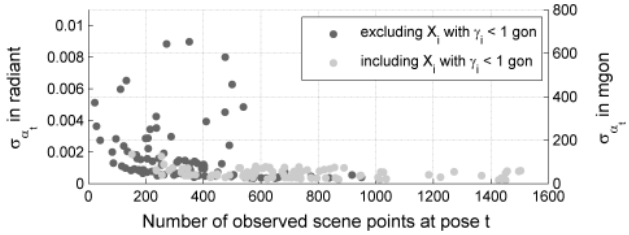
Software. Matlab code of the proposed bundle adjustment BACS is avaible at: `www.ipb.uni-bonn.de/bacs`.



(a) Number of scene points with small intersection angles.

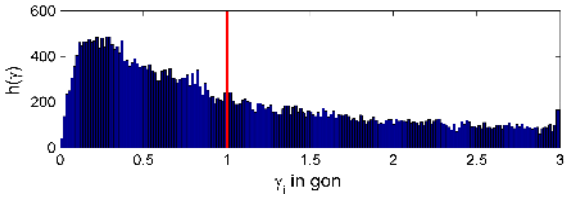


(b) Average empirical standard deviation of estimated rotations.

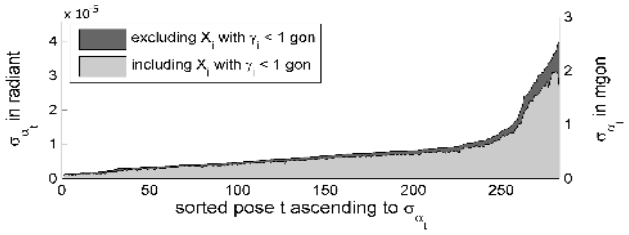


(c) Scatter plot of σ_{α_t} against the number of observed scene points at t.

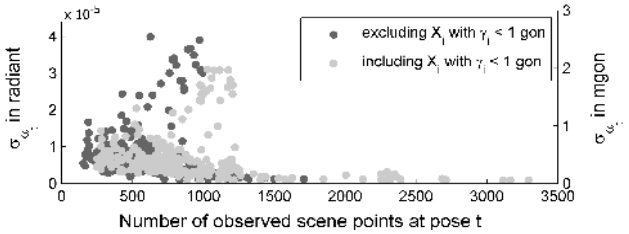
Figure 7: The histogram in (a) shows the number of scene points in the multi-camera dataset with small intersection angles. The average precision σ_{α_t} determined by excluding and including scene points with $\gamma < 1$ gon for all poses $t = 1, ..., T$ is compared to each other in (b) and against the number of observed scene points in (c).



(a) Number of scene points with small intersection angles.



(b) Average empirical standard deviation of estimated rotations.



(c) Scatter plot of σ_{α_t} against the number of observed scene points at t.

Figure 8: The subplots (a), (b) and (c) show the results of the investigations explained in Fig. 7 on the urban drive dataset.

Acknowledgements. This work was partially supported by the DFG-Project FOR 1505 "Mapping on Demand". The authors wish to thank the anonymous reviewers for their helpful comments.

References

Bartoli, A., 2002. On the Non-linear Optimization of Projective Motion Using Minimal Parameters. In: Proceedings of the 7th European Conference on Computer Vision-Part II, ECCV '02, pp. 340–354.

Förstner, W., 2012. Minimal Representations for Uncertainty and Estimation in Projective Spaces. Zeitschrift für Photogrammetrie, Fernerkundung und Geoinformation, Vol. 3, pp. 209–220.

Frahm, J.-M., Köser, K. and Koch, R., 2004. Pose estimation for Multi-Camera Systems. In: Mustererkennung '04, Springer.

Hartley, R. I. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.

Heuel, S., 2004. Uncertain Projective Geometry: Statistical Reasoning for Polyhedral Object Reconstruction. LNCS, Vol. 3008, Springer. PhD. Thesis.

Huber, P. J., 1981. Robust Statistics. John Wiley, New York.

Kanatani, K., 1996. Statistical Optimization for Geometric Computation: Theory and Practice. Elsevier Science.

Kim, J.-H., 2010. Camera Motion Estimation for Multi-Camera Systems. PhD thesis, School of Engineering, ANU College of Engineering and Computer Science, The Australian National University.

Kraus, K., 1997. Photogrammetry. Dümmler Verlag Bonn. Vol.1: Fundamentals and Standard Processes. Vol.2: Advanced Methods and Applications.

Läbe, T. and Förstner, W., 2006. Automatic Relative Orientation of Images. In: Proceedings of the 5th Turkish-German Joint Geodetic Days.

Lourakis, M. I. A. and Argyros, A. A., 2009. SBA: A Software Package for Generic Sparse Bundle Adjustment. ACM Transactions on Mathematical Software 36(1), pp. 1–30.

McGlone, C. J., Mikhail, E. M. and Bethel, J. S., 2004. Manual of Photogrammetry. American Society of Photogrammetry and Remote Sensing.

Meidow, J., Beder, C. and Förstner, W., 2009. Reasoning with Uncertain Points, Straight Lines, and Straight Line Segments in 2D. International Journal of Photogrammetry and Remote Sensing 64, pp. 125–139.

Mostafa, M. M. R. and Schwarz, K. P., 2001. Digital Image Georeferencing from a Multiple Camera System by GPS/INS. PandRS 56(1), pp. 1–12.

Mouragnon, E., Lhuillier, M., Dhome, M. and Dekeyser, F., 2009. Generic and Real-time Structure from Motion Using Local Bundle Adjustment. Image and Vision Computing 27(8), pp. 1178–1193.

Muhle, D., Abraham, S., Heipke, C. and Wiggenhagen, M., 2011. Estimating the Mutual Orientation in a Multi-Camera System with a Non Overlapping Field of View. In: Proceedings of the 2011 ISPRS Conference on Photogrammetric Image Analysis.

Nister, D., Naroditsky, O. and Bergen, J., 2004. Visual Odometry. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 652–659.

Savopol, F., Chapman, M. and Boulianne, M., 2000. A Digital Multi CCD Camera System for Near Real-Time Mapping. In: Intl. Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Proc. ISPRS Congress, Amsterdam.

Schmeing, B., Läbe, T. and Förstner, W., 2011. Trajectory Reconstruction Using Long Sequences of Digital Images From an Omnidirectional Camera. In: Proceedings of the 31th. DGPF Conference.

Zomet, A., Wolf, L. and Shashua, A., 2001. Omni-rig: Linear Self-recalibration of a Rig with Varying Internal and External Parameters. In: Proc. Eighth IEEE Int. Conf. on Computer Vision, pp. 135–141.