A HYBRID GLOBAL IMAGE ORIENTATION METHOD FOR SIMULTANEOUSLY ESTIMATING GLOBAL ROTATIONS AND GLOBAL TRANSLATIONS

Xin Wang[1, *], Teng Xiao[2], Yoni Kasten[3]

[1]Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany, wang@ipi.uni-hannover.de
[2]School of Geodesy and Geomatics, Wuhan University, Wuhan, PR.China, xiaoteng@whu.edu.cn
[3]Weizmann Institute of Science, Israel, yoni.kasten@weizmann.ac.il

Commission II, WG II/1

KEY WORDS: image orientation, global structure from motion (SfM), global rotations estimation, global translation estimation

ABSTRACT:

In recent years, the determination of global image orientation, i.e. global SfM, has gained a lot of attentions from researchers, mainly due to its time efficiency. Most of the global methods take relative rotations and translations as input for a two-step strategy comprised of global rotation averaging and global translation averaging. This paper by contrast presents a hybrid approach that aims to solve global rotations and translations simultaneously, but hierarchically. We first extract an optimal minimum cover connected image triplet set (OMCTS) which includes all available images with a minimum number of triplets, all of them with the three related relative orientations being compatible to each other. For non-collinear triplets in the OMCTS, we introduce some basic characterizations of the corresponding essential matrices and solve for the image pose parameters by averaging the constrained essential matrices. For the collinear triplets, on the other hand, the image pose parameters are estimated by relative orientation using the depth of object points from individual local spatial intersection. Finally, all image orientations are estimated in a common coordinate frame by traversing every solved triplet using a similarity transformation. We show results of our method on different benchmarks and demonstrate the performance and capability of the proposed approach by comparing with other global SfM methods.

1. INTRODUCTION

Image orientation (also known as Structure-from-Motion - SfM or pose estimation) plays a key role in the field of photogrammetry and computer vision. Although this topic has been very well studied in the last several decades, it recently again caught the interest of photogrammetrists due to the increasing number of images (e.g., images shared through websites) and images taken without proper acquisition planning. Today, according to the procedure in which images are oriented, there are typically three different strategies to solve this problem: incremental, hierarchical and global methods. Incremental SfM (Snavely et al., 2006; Agarwal et al., 2009; Schönberger and Frahm, 2016; Wu, 2013; Wang et al, 2018 and 2019a) starts with an initial subset of images, e.g., initializing a small reconstruction, and iteratively adds further images to the block with repetitive intermediate bundle adjustment. Farenzena, et al. (2009), Mayer (2014) and Toldo, et al. (2015) present a so called hierarchical method, which improves the incremental idea by first dividing the images into overlapping subsets, and then processing all subsets individually by incremental SfM, finally merging them in a hierarchical way with a number of bundle adjustments. Both of these strategies are relatively slow because of the repeated use of bundle adjustments. To overcome this problem, Martinec & Pajdla (2007), Arie-Nachimson et al (2012), Jiang et al. (2013), Moulon et al. (2013), and Wang et al. (2019a and 2019b) present global SfM methods which first estimates all available image pose parameters, and then perform only one final bundle adjustment for refinement. All above mentioned global methods have a common limitation: they work in two individual steps. Only after image rotations have been solved, the translation parameters can be estimated. This can create problems, if rotations are incorrectly estimated. In addition, many researchers (Wilson and Snavely, 2014; Shah et al, 2018) also found that global methods are sensitive to outliers of relative orientations, since outliers are difficult to detect in global computations.

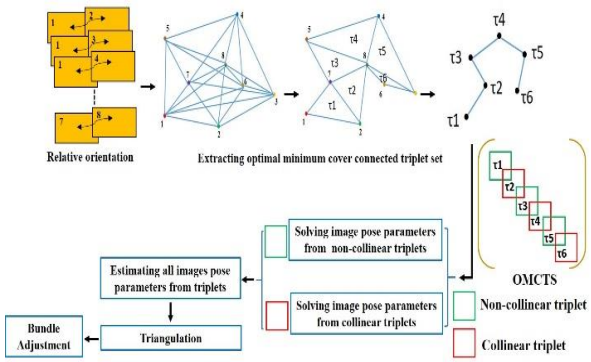


Figure 1. The workflow of our hybrid image orientation method, where in this figure τ_i denotes the i-th selected triplet.

We are most interested in those time efficient strategies, and thus present a novel hybrid global image orientation approach. To improve the time efficiency and robustness, among the overlapping image pairs and their corresponding relative orientations we first extract an optimal minimum cover connected triplet set (OMCTS) such that it not only includes all available images with a minimum number of triplets, but also makes the corresponding relative orientations within extracted triplets as compatible as possible. Then, we apply a hybrid method by considering non-collinear and collinear triplets separately, where *collinear* means the three image projection centres are collinear (see Fig. 1 for the workflow of our method). For the non-collinear triplets, we make use of algebraic constraints of the corresponding essential matrices and derive eligible essential matrices, subsequently image pose parameters are computed by essential matrix averaging. For collinear triplets, this method is invalid, thus their image pose parameters are recovered by using relative orientations with the depth of object points from individual local spatial intersection. As the estimated

* Corresponding author

image pose parameters are given in local coordinate systems of each individual triplet, we need to transform them into a global unified system. We do so by traversing image poses of each individual triplet using similarity transformations. Finally, we run one single bundle adjustment to refine our results.

Our main contributions are *threefold*: *First*, we present an idea to build an optimal minimum cover connected triplet set that can combine both, collinear and non-collinear triplets. *Second*, we introduce a hybrid method to solve the image pose parameters for both non-collinear and collinear triplets separately, from which global pose parameters in a unified coordinate system are estimated. *Finally*, via testing various settings for solving non-collinear triplets, a very reasonable set is suggested, and our hybrid method is then evaluated by comparing the results with other global SfM methods using various benchmarks.

2. RELATED WORK

Recently, research of image pose estimation or SfM has become very active, since more and more complicated datasets have become available, e.g. images downloaded from the Internet and images with repetitive structures and critical configurations (Wang et al., 2019c). In this section we review some state-of-the-art works in global SfM research. Typically, global SfM methods are conducted in two separate steps, global rotation estimation and the subsequent global translation estimation. However, there are also some works, called integrative global methods, which estimate global rotations and translations simultaneously.

Global rotation estimation. The problem of global rotation estimation from relative rotations of image pairs have been studied by many researchers. Govindu (2001) use quaternions to represent rotations, global quaternions are determined by a constrained least squares optimization. Martinec and Pajdla (2007), Arie-Nachimson et al (2012) and Moulon et al. (2013) first relax the constraints on rotation parameters and present a linear homogeneous equation system, which is solved using SVD (singular value decomposition). Hartley et al. (2011) present a robust iterative method using L1 norm optimization based on the Lie algebra of *SO*(3). Chatterjee and Govindu (2013) present a two-stage approach: they first calculate initial global rotation using a minimum spanning tree and then refine the solution with an iterative reweighting scheme combining the Lie algebra of *SO*(3). Reich et al. (2016, 2017) present a method which extends the approach of Chatterjee and Govindu (2013), they studied the algebraic characterization of relative rotations in multi-image settings and apply a convex relaxed semidefinite program to obtain a more robust initial solution which is further refined by using Lie algebra of *SO*(3).

Global translation estimation. Unlike global rotation parameters, global translation parameters cannot be directly estimated, since the baseline of an image pair has an arbitrary length. Nevertheless, many methods have been studied for global translation estimation. Govindu (2001) present an iterative reweighting scheme to obtain global scale unified translation vectors; however, this method is invalid in degenerate cases, e.g., when projection centres of images are (nearly) collinear. Jiang et al. (2013) present a solution which can solve degenerate cases by using depth information of tie points; a global linear equation system is built by concatenating connected triplets. Since the triplets are required to be well connected, this method normally recovers fewer images. It was extended by Cui et al. (2015) and Wang et al. (2019b), they first solve for the global scale factor for each eligible relative translation and then resize all relative translations such that they are all in the same global scale unified system, and finally global translations are estimated by using those resized relative translations. Wilson and Snavely (2014) propose a method called 1DSfM, to robustify their result, they first detected blunders of relative translations by projecting the 3D relative translation into different 1D direction vectors. Typically, the blunders clearly stand out in some directions of the 1D vectors. Then, a non-linear method based on inliers of relative translations and tie points is proposed, this non-linear method is not guaranteed to converge when outliers exist. By using collinearity equations and the information of tie points, Wang et al. (2019a) propose a linear global method. Given the global rotation and tie point information, they first selected some robust tie points that can connect all available images into the same photogrammetric block. Then, the translation parameters and selected 3D tie point coordinates are solved simultaneously. But, as the number of images increases, so does the number of unknown tie points, which brings much more computational burden for the linear global method.

Integrative global method. Recently, ideas were published to avoid having to compute rotation and translation separately. Bourmaud et al. (2014) derive the image pose parameters as a Lie group *SE(3)*, they propose a generative model based on the formulation of a concentrated Gaussian distribution on the matrix Lie group and solve an iterated extended Kalman filter on that group to compute the elements of *SE(3)*. Kasten et al. (2019a) propose a method to globally recover the projection matrix of each image by using fundamental matrices of image pairs. However, as the projection matrix yields a projective reconstruction, information on interior orientation parameters cannot be introduced. Later, the authors extended their work. Exploring the algebraic characterizations of essential matrices, they introduced a method to simultaneously solve for rotation and translation of each image from essential matrices (Kasten et al., 2019b). The disadvantage is that this method cannot deal with projection centres that are all (nearly) collinear.

The remainder of this paper is structured as follows: In Section 3 we introduce some basics of essential matrices in multi-image settings. Section 4 describes our method of estimating image pose parameters by using the information of triplets. In Section 5, we report results of experiments on various benchmarks to evaluate our method. Finally, Section 6 concludes our work.

3. THE N-IMAGE ESSENTIAL MATRIX

Following partial content of Kasten et al. (2019b) to make this paper more self-contained, we next give some definitions and corollaries with respect to the so called N-Image essential matrix. Given a set of n images which are denoted as $1, 2, 3, \ldots, n$, let $t_i \in \mathbb{R}^3$ and $R_i \in SO(3)$ be the translation and rotation parameters of image i in a global coordinate system. The essential matrix of two images i and j can be derived as $E_{ij}=R_i^T(T_i-T_j)R_j$, where $T_i=[t_i]_\times$ is the skew-symmetric matrix of vector t_i.

Definition 1. A matrix $E \in Sym_{3n}$ (Sym_{3n} denotes the space of all the 3n×3n symmetric matrices), whose 3×3 block matrices are denoted by E_{ij}, is called a N-Image essential matrix if $\forall\ i \neq j$, $rank(E_{ij})=2$, and the corresponding two eigenvalues are equal, $\forall\ E_{ii}=\mathbf{0}$, where $\mathbf{0}$ denotes the corresponding zero matrix.

Corollary 1. A N-Image essential matrix E is *scale consistent*, if there exist n rotation matrices $\{R_i\}_{i=1,\ldots,n}$, n projection centres $\{t_i\}_{i=1,\ldots,n}$ and n non-zero scalars $\{\alpha_i\}_{i=1,\ldots,n}$ such that $E_{ij}=\alpha_i R_i^T(T_i-T_j)R_j\alpha_j$. Given the constraint that not all projection centres of $\{t_i\}_{i=1,\ldots,n}$ are collinear, the SVD of E can be then derived as $E=[\hat{M}\ \hat{N}]\begin{bmatrix}\Sigma_+ & \\ & \Sigma_+\end{bmatrix}\begin{bmatrix}\hat{M}^T \\ \hat{N}^T\end{bmatrix}$ and $rank(E)=6$, where $\hat{M}, \hat{N} \in \mathbb{R}^{3n\times 3}$ and $\Sigma_+ \in$

$\mathbb{R}_+^{3\times3}$. What's more, the following conditions are sufficient and necessary conditions: *First*, E is a scale consistent N-Image essential matrix; *Second*, The SVD of E can be written as $E=[\widehat{M}\ \widehat{N}]\begin{bmatrix}\Sigma_+ & \\ & \Sigma_+\end{bmatrix}\begin{bmatrix}\widehat{M}^T\\ \widehat{N}^T\end{bmatrix}$ with $\widehat{M}$, $\widehat{N} \in \mathbb{R}^{3n\times3}$ and $\Sigma_+ \in \mathbb{R}_+^{3\times3}$, such that each block of $\widehat{N}$ denoted as $\widehat{N}_i$, is a scaled rotation matrix, i.e., $\widehat{N}_i=\hat{\alpha}_i\hat{R}_i$ and $\hat{R}_i \epsilon SO(3)$. $\widehat{N}$ is called scaled block rotation matrix; *Third*, $\Sigma_+=-\Sigma_-$ and the spectral decomposition of E reads $E=[A\ B]\begin{bmatrix}\Sigma_+ & \\ & \Sigma_-\end{bmatrix}\begin{bmatrix}A^T\\ B^T\end{bmatrix}$ with $A, B \in \mathbb{R}^{3n\times3}$, $\Sigma_+=-\Sigma_-$ and $\Sigma_+, \Sigma_- \in \mathbb{R}_+^{3\times3}$, where $\sqrt{0.5}\,(A+B)$ is a scaled block rotation matrix (see the corresponding proofs in Kasten et al. (2019b)).

A scale consistent N-image essential matrix thus is a matrix representation of the orientation parameters of the N images, coding them in a similar way that the essential matrix does for two images, so that rays of conjugate points intersect.

Corollary 2. Again following Kasten et al. (2019b), it is possible to determinate all image rotation matrices $\{R_i\}_{i=1,\dots,n}$, projection centres $\{t_i\}_{i=1,\dots,n}$ (in a global coordinate system) and n non-zero scalars$\{\alpha_i\}_{i=1,\dots,n}$, given a scale consistent N-Image essential matrix E, where the camera projection centres are not all collinear, in the following way.

1. Do spectral decomposition of E and obtain the eigenvectors A, B of E together with the corresponding eigenvalues to be found in Σ_+ and Σ_-. SVD decomposition is not used, because a standard SVD method has multiplicity of singular values on E with the corresponding rank being equal to 6 and typically sorts the singular values in a descending order, which doesn't produce the specific SVD form as corollary 1 explains.

2. There are in total eight possibilities of $\sqrt{0.5}(A+BI_t)$ with $I_t = \begin{pmatrix}\pm1 & 0 & 0\\ 0 & \pm1 & 0\\ 0 & 0 & \pm1\end{pmatrix}$, because of the sign ambiguity of each eigenvector which can be solved by equation (6), see below.

3. $\widehat{N}=\sqrt{0.5}\,(A+BI_t)$, the scalar of each block $\widehat{N}_i$ can be computed by $\hat{\alpha}_i=(\det(\widehat{N}_i))^{1/3}$, and $R_i=(\widehat{N}_i/\hat{\alpha}_i)^{\mathrm{T}}$.

4. $E_{ij}=\widehat{M}_i\Sigma_+\widehat{N}_j^{\,T}+\widehat{N}_i\Sigma_+\widehat{M}_j^{\,T}$, as $E_{ii}=\mathbf{0}$ we see that $\widehat{M}_i\Sigma_+\widehat{N}_i^{\,T}$ is skew symmetric; we can derive the projection centre $[t_i]_\times=\widehat{N}_i^{-1}\widehat{M}_i\Sigma_+$.

The N-Image essential matrix can thus be regarded as a tool to estimate rotations and translations simultaneously from pairwise essential matrices. However, three practical difficulties exist: *First*, we can't compute every essential matrix for each pair, because many image pairs do not overlap; *second*, calculated essential matrices are typically normalized, e.g., when employing the 5-Point algorithm (Nistér, 2004), thus it is very difficult to guarantee for a N-Image essential matrix to be scale consistent if N>3, because the non-zero scalars cannot be set arbitrarily; *third*, the case that all projection centres are (or nearly are) collinear does exist in many applications, e.g., images captured by mobile mapping car moving along a straight line or aerial images within one strip.

4. METHODOLOGY

To solve these three practical difficulties, we investigate triplets instead of larger sets of images, which overcome the first two points and then present a hybrid method to separately deal with collinear and non-collinear triplets to avoid the third difficulty.

We first introduce corollary 3.

Corollary 3. Given a non-collinear triplet, the corresponding scale consistent 3-Image essential matrix is invariant to scales (see our proof in the appendix).

4.1 Generation of an optimal minimum cover connected image triplet set

We use three images with mutual overlap and extract all related triplets, a corresponding triplet graph is then built as Fig. 1 shows: triplets denote nodes and two triplets are connected to each other, if they share two common images. An optimal subset of these triplets is selected for better time efficiency and robustness. We select such a subset called optimal minimum cover connected image triplet set (OMCTS) with the following requirements: 1) the selected triplets cover all available images and the three relative orientations should be as compatible as possible; 2) triplets from the selected subset are connected, which guarantees that the photogrammetric block will not break; 3) the minimum number of triplets that fulfil the above two requirements is selected.

To identify the compatibility of each triplet, similar to Wang et al. (2019b) and Kasten et al. (2019a), we compute two triplet closure discrepancies with respect to relative rotations and translations, respectively. Given three relative rotations of a triplet, R_{ij}, R_{jk} and R_{ki}, $R_{ij}R_{jk}R_{ki} = I_{3\times3}$ should hold. However, this is not strictly the case because of outliers and noise in relative rotations. We can use $d_\angle(s_R, I_{3\times3}) = \arccos((tr(R_{ij}R_{jk}R_{ki} - I_{3\times3}) - 1)/2)$ as one indicator of the triplet compatibility. The discrepancy in relative translation can be calculated from the difference of the sum of the angles formed by the three projection centres within a triplet and 180°, i.e., $d_\angle(s_T, 180°) = |\theta_i + \theta_j + \theta_k - 180°|$ with $\theta_i = \arccos\frac{t_{ij}t_{ik}}{||t_{ij}||||t_{ik}||}$ and ||.|| the L2 norm (see Wang et al., 2019b for details). Based on these two criteria, the triplet compatibility indicator is formulated as $max\ (d_\angle(s_R, I_{3\times3}), d_\angle(s_T, 180°))$. Finally, we employ a greedy triplet deleting scheme: starting with the triplet with the largest indicator, a triplet is deleted as long as the remaining triplets are still connected and no image is deleted from the photogrammetric block (see Appendix for more details on generating the OMCTS), note that we introduce our triplet selection process in a less sophisticated way and a more grounded graph theory based explanation is given by Shah et al. (2018).

The collinearity degree of a triplet is determined by the minimal angle among θ_i, θ_j and θ_k. From the triplets selected for the OMCTS, the ones with that minimal angle larger than a threshold θ_{ang} are considered to be non-collinear, the others are considered collinear.

4.2 Solving image pose for non-collinear triplets

Based on Kasten et al. (2019b), this section focuses on non-collinear triplets which are denoted as $\{\tau_{nc}\}_{nc=1}^K$, K is the number of detected non-collinear triplets and *nc* is the *nc-th* non-collinear triplet, the corresponding 3-Image essential matrix is denoted as $\{E_{\tau_{nc}}\}_{nc=1}^K$, the elements of $\{E_{\tau_{nc}}\}_{nc=1}^K$ are the unknowns. As input, we have corresponding estimated essential matrices $\check{E}_{ij}$(e.g., using the 5-point algorithm) for each overlapping image pair, and they can be transformed into estimated 3-Image essential matrices denoted as $\{\check{E}_{\tau_{nc}}\}_{nc=1}^K$.

Our goal is to first seek a scale consistent 3-Image essential matrix that is as close as possible to the estimated 3-Image essential matrix for all non-collinear triplets and then estimate exterior pose parameters within each non-collinear triplet by using corollary 2. The constrained problem can be formulated as

$$\underset{\{E_{\tau_{nc}}\}_{nc=1}^K}{\text{minimize}} \sum_{nc=1}^K ||E_{\tau_{nc}} - \check{E}_{\tau_{nc}}||_F^2 \quad (1)$$

subject to $rank(E_{\tau_{nc}}) = 6$; $\Sigma_+(E_{\tau_{nc}}) = -\Sigma_-(E_{\tau_{nc}})$; $\sqrt{0.5}(A(E_{\tau_{nc}}) + B(E_{\tau_{nc}}))$ is a block rotation,

where $\Sigma_+(E_{\tau_{nc}})$, $-\Sigma_-(E_{\tau_{nc}})$ are the 3 largest eigenvalues in descending order and the 3 smallest eigenvalues in ascending order of $E_{\tau_{nc}}$, respectively. $A(E_{\tau_{nc}})$ and $B(E_{\tau_{nc}})$ are the corresponding eigenvectors. Solving (1) is not easy due to the non-convex rank defect and block rotation constraints. The the alternating direction method of multipliers (ADMM, Boyd et al., 2011) is used to solve equation (1) iteratively; we can generate an equivalent constrained optimization problem.

$$\max_{\vartheta,\rho} \min_{\{E_{\tau_{nc}}\},W,Q} \sum_{nc=1}^{K} l(\{E_{\tau_{nc}}\}, W_{nc}, \vartheta_{nc}, Q_{nc}, \rho_{nc}) \quad (2)$$

subject to $rank\ (W_{nc}) = 6$; $\Sigma_+(W_{nc}) = -\Sigma_-(W_{nc})$; $rank\ (Q_{nc}) = 6$; $A(Q_{nc}) + B(Q_{nc})$ is a block rotation.

where, $l(\{E_{\tau_{nc}}\}, W_{nc}, \vartheta_{nc}, Q_{nc}, \rho_{nc}) = ||E_{\tau_{nc}} - \check{E}_{\tau_{nc}}||_F^2 + \Delta_1||W_{nc} - E_{\tau_{nc}} + \vartheta_{nc}||_F^2 + \Delta_2||Q_{nc} - E_{\tau_{nc}} + \rho_{nc}||_F^2$, W_{nc} and Q_{nc} are auxiliary matrices for constraints of rank defect and block rotation, respectively. ϑ_{nc} and ρ_{nc} are two Lagrange multipliers. Initializations are given at $sp = 0$ (sp denotes the number of iterations) as ${W_{nc}}^0 = {Q_{nc}}^0 = \check{E}_{\tau_{nc}}$, ${\vartheta_{nc}}^0 = {\rho_{nc}}^0 = 0$. We then solve (1) iteratively by alternating between the following steps:

(a) *Computing* $\{E_{\tau_{nc}}\}$

$$\{E_{\tau_{nc}}\}^{sp} = \underset{\{E_{\tau_{nc}}\}}{\operatorname{argmin}} \sum_{nc=1}^{k} ||E_{\tau_{nc}} - \check{E}_{\tau_{nc}}||_F^2 + \Delta_1||W_{nc} - E_{\tau_{nc}} + \vartheta_{nc}||_F^2 + \Delta_2||Q_{nc} - E_{\tau_{nc}} + \rho_{nc}||_F^2$$

This is a convex quadratic optimization problem and can be solved using equation (3)

$${E_{\tau_{nc}}}^{sp} = \frac{1}{1+2\Delta_1+2\Delta_2}[\ 2\Delta_2(\ {Q_{nc}}^{sp-1} + {\rho_{nc}}^{sp-1}) + 2\Delta_1(\ {W_{nc}}^{sp-1} + {\vartheta_{nc}}^{sp-1}) + \check{E}_{\tau_{nc}}\ ] \quad (3)$$

(b) *Computing* W_{nc}

$${W_{nc}}^{sp} = \underset{W_{nc}}{\operatorname{argmin}} ||W_{nc} - {E_{\tau_{nc}}}^{sp} + {\vartheta_{nc}}^{sp-1}||_F^2 \quad (4)$$

subject to $rank\ (W_{nc}) = 6$; $\Sigma_+(W_{nc}) = -\Sigma_-(W_{nc})$;

Equation (4) is also convex quadratic optimization problem, thus, the estimation of ${W_{nc}}^{sp}$ should be ${E_{\tau_{nc}}}^{sp} - {\vartheta_{nc}}^{sp-1}$, however, this may not fulfil the constraints of rank defect and eigenvalue of $\Sigma_+(E_{\tau_{nc}}) = -\Sigma_-(E_{\tau_{nc}})$. To overcome this issue, we do a spectral decomposition on ${E_{\tau_{nc}}}^{sp} - {\vartheta_{nc}}^{sp-1}$ by $\overline{\overline{U}}\Sigma'\overline{\overline{U}}^T$, $\overline{\overline{U}}$ is a 9×9 matrix and Σ' a diagonal matrix with corresponding eigenvalues sorted in descending order. Thus, we can update W_{nc} as

$${W_{nc}}^{sp} = \overline{\overline{U}}\Sigma^*\overline{\overline{U}}^T \quad (5)$$

where $\Sigma_{11}^* = 0.5(\Sigma_{11}' - \Sigma_{99}')$, $\Sigma_{22}^* = 0.5(\Sigma_{22}' - \Sigma_{88}')$, $\Sigma_{33}^* = 0.5(\Sigma_{33}' - \Sigma_{77}')$, $\Sigma_{44}^* = 0$, $\Sigma_{55}^* = 0$, $\Sigma_{66}^* = 0$, $\Sigma_{77}^* = 0.5(\Sigma_{77}' - \Sigma_{33}')$, $\Sigma_{88}^* = 0.5(\Sigma_{88}' - \Sigma_{22}')$, $\Sigma_{99}^* = 0.5(\Sigma_{99}' - \Sigma_{11}')$.

(c) *Computing* Q_{nc}

$${Q_{nc}}^{sp} = \underset{Q_{nc}}{\operatorname{argmin}} ||Q_{nc} - {E_{\tau_{nc}}}^{sp} + {\rho_{nc}}^{sp-1}||_F^2 \quad (6)$$

subject to $rank\ (\rho_{nc}) = 6$; $A(Q_{nc}) + B(Q_{nc})$ is a block rotation.

Similar to equation (4), the initial guess of ${Q_{nc}}^{sp}$ would be ${E_{\tau_{nc}}}^{sp} - {\rho_{nc}}^{sp-1}$, which may violate the extra constraints. To obtain an eligible solution, we do a spectral decomposition for the initial guess. Σ_+ and Σ_- are eigenvalues, A and B are corresponding eigenvectors, A, $B \in \mathbb{R}^{9\times9}$. Now, the requirement is that $\sqrt{0.5}\ (A + BI_t)$ with $I_t = \begin{pmatrix} \pm1 & 0 & 0 \\ 0 & \pm1 & 0 \\ 0 & 0 & \pm1 \end{pmatrix}$ is a block rotation matrix. To find a correct solution for I_t, we have ${I_t}^* = \underset{I_t}{\operatorname{argmax}} \sum_{i=1}^{3} \frac{||diag(\sqrt{0.5}\ (A_i+B_iI_t)^T\sqrt{0.5}\ (A_i+B_iI_t))||_2}{||\sqrt{0.5}\ (A_i+B_iI_t)^T\sqrt{0.5}\ (A_i+B_iI_t)||_F}$, A_i and B_i is the corresponding block matrix of A and B. This is also applied in corollary 2.

Let $\ddot{N} = [\ddot{N}_1 \quad \ddot{N}_2 \quad \ddot{N}_3]^T$, where $\ddot{N}_i$ is the closest scaled rotation of $\sqrt{0.5}\ (A_i + B_iI_t^*)$, which is obtained by first computing a SVD of $\sqrt{0.5}\ (A_i + B_iI_t^*)$ and replacing the diagonal matrix of singular values by an 3×3 identity matrix, the average of original singular values is the scale factor. Let $\ddot{M} = \sqrt{0.5}\ (A_i - B_iI_t^*)$, we update A and B by $\ddot{A} = \sqrt{0.5}\ (\ddot{M} + \ddot{N})$ and $\ddot{B} = \sqrt{0.5}\ (\ddot{N} - \ddot{M})$, finally

$$Q_{nc}{}^{sp} = [\ddot{A}\ \ddot{B}] \begin{bmatrix} \Sigma_+ & \\ & \Sigma_- \end{bmatrix} \begin{bmatrix} \ddot{A}^T \\ \ddot{B}^T \end{bmatrix} \quad (7)$$

(d) *Computing* ϑ_{nc} and ρ_{nc}

$${\vartheta_{nc}}^{sp} = {\vartheta_{nc}}^{sp-1} + {W_{nc}}^{sp} - {E_{\tau_{nc}}}^{sp} \quad (8)$$

$${\rho_{nc}}^{sp} = {\rho_{nc}}^{sp-1} + {Q_{nc}}^{sp} - {E_{\tau_{nc}}}^{sp} \quad (9)$$

In our experiments, we set $\Delta_1 = 100$ and $\Delta_2 = 0.01$ to weight rank defect and block rotation constraints, respectively, and repeat the above four steps 100 times (more interpretations related to settings of Δ_1, Δ_2 and sp are discussed in our experimental section below) and we obtain the scale consistent 3-Image essential matrix. Rotation and translation of each image within one non-collinear triplet are then estimated using corollary 2.

4.3 Solving image pose parameters from collinear triplets

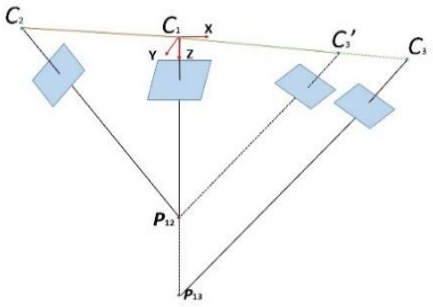


Figure 2. Collinear triplet Case.

Different to Kasten et al. (2019b), where the authors deleted all (nearly) collinear triplets, both non-collinear and collinear triplets are considered in this paper. To deal with collinear triplets, we choose one image as reference and use the information of relative rotations and translations to estimate the exterior orientation parameters of the other two images. Global rotations within one triplet are straightforward to compute: we assign an identity matrix to one image and obtain the other two rotations by propagating the relative rotations. However, global translations within one triplet are not that easy to compute, because the length of relative translations are typically normalized to 1 when decomposing the essential matrix, and this will normally lead to scale ambiguity as Fig. 2 shows. The projection centres C_1, C_2 and C_3 of images {*1*, *2*, *3*} are collinear, which generates a collinear triplet. P_{12} and P_{13} represent the same object point, but have different positions after triangulation due to the different scales of the two models. Fig. 2 implies that we can remove the scale ambiguity by moving the original C_3 to C_3', mathematically this can be expressed by using the depth values of calculated position of P_{12} and P_{13}. We have

$$|C_1C_3'|/|C_1C_3| = |C_1P_{12}|/|C_1P_{13}| = Z_{P_{12}}/Z_{P_{13}} = \lambda \quad (10)$$

where $|.|$ returns length, $Z_{P_{12}}$ and $Z_{P_{13}}$ are the corresponding Z values (as object points are always in front of cameras, the Z value is guaranteed to be larger than 0). Each three-ray point contributes one λ, we use the idea of Wang et al. (2019b) to obtain

a robust solution $\dot{\lambda}$. Given $\dot{\lambda}$ and the relative rotations R_{12}, R_{13} and relative translation t_{12}, t_{13}, we obtain a triplet of scale consistent exterior orientation parameters by formula (11) with the assumption that image pair (i_1,i_2) has most correspondences within the image pairs of the triplet (note that R_{23} and t_{23} are not used in this solution to reduce the computational complexity, and we assume that the relative orientations within the selected compatible triplets can be considered to be accurate after having checked them before).

$$\begin{aligned} R_1 &= I_{3\times3} \quad t_1 = \mathbf{0} \\ R_2 &= R_{12} \;\; t_2 = t_{12} \\ R_3 &= R_{13}, t_3 = \dot{\lambda} \cdot t_{13} \end{aligned} \tag{11}$$

For all detected collinear triplets, equation (11) is used to obtain rotation and translation of each image.

4.4 Estimating all images pose parameters from triplets

We have now estimated the exterior orientation parameters (three rotations and translations per image) within all triplets, whether collinear or non-collinear, which are uniquely determined up to a similarity transformation. For any two connected triplets which share two common images, there is a possibility to compute a unique similarity transformation between these two triplets by using the two corresponding common image pose parameters calculated from individual triplet (Hartley and Zisserman, 2004). Since a minimum cover connected image triplet set has already been generated and the corresponding pose parameters within the triplets are available, the extracted connected triplets can be traversed and similarity transformations between all connected triplets can be applied to transform all exterior orientation parameters into a common coordinate system (see the Appendix for more details of calculating the similarity transformation between two connected triplets).

5. EXPERIMENTS

To evaluate our method, we implemented the proposed global hybrid image orientation method as the workflow in Fig. 1 shows. We set the free parameter θ_{ang} to be 0.17 (in radian) for all experiments [1]. The experiments are first conducted on four terrestrial close range datasets, one of them is a public dataset with 128 images around a *building* (Zach, et al. 2010) which consists of both (nearly) collinear and non-collinear images. The other three test data are benchmark datasets published by Strecha et al. (2008) which are made up of 11 to 30 images. Each of these three datasets is provided with ground truth exterior orientation parameters, which are used for comparison. Finally, we further explore our method by dealing with one set of oblique quasi-aerial images from an open public photogrammetric contest [2] (Özdemir et al., 2019). The bundle adjustment of Wang et al. (2019b) integrated with the open source Ceres-solver (Agarwal et al., 2017) is applied for refining the results.

5.1 Analyzing various settings of Δ_1, Δ_2 and *sp*

To inspect the influence of Δ_1, Δ_2 and *sp* on solving equation (1), we first investigate the rank constraints (i.e., $rank(E_{\tau_{nc}}) = 6$) on *castle-P30* by calculating the logarithm of the mean ratio between the 7-*th* and 6-*th* singular values $log_{10}(\sigma_7/\sigma_6)$ of all triplets in $\{E_{\tau_{nc}}\}_{nc=1}^{K}$ for different settings of Δ_1, Δ_2 and *sp*. In general, a reliable solution of a 3-Image scale consistent essential matrix from equation (1) can generate a very small value for $log_{10}(\sigma_7/\sigma_6)$. The results shown in Fig. 3 indicate that in our experiment $log_{10}(\sigma_7/\sigma_6)$ decreases as the iteration process runs and it starts to become stable at the 80-*th* iteration. The case of $\Delta_1 > \Delta_2$ normally generates much smaller values for $log_{10}(\sigma_7/\sigma_6)$ than that of $\Delta_1 \leq \Delta_2$ does. Also, the larger the ratio of Δ_1 / Δ_2 is, the smaller $log_{10}(\sigma_7/\sigma_6)$ becomes in general, because only if the rank constraint is fulfilled, can the spectral decomposition be processed for the block rotation constraint. So, we typically set a high weight Δ_1.

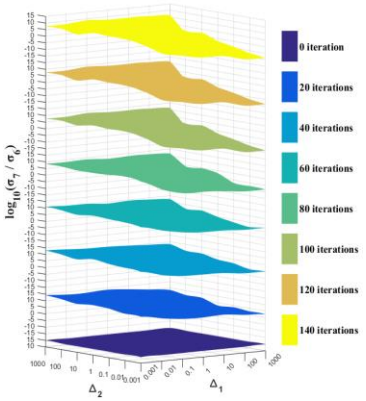


Figure 3. Rank constraints of various setting of Δ_1, Δ_2 and *sp*.

However, as Fig. 3 shows, we can't conclude that an infinitely large Δ_1 is best, because this will lead to the constraint that $\sqrt{0.5}(A(E_{\tau_{nc}}) + B(E_{\tau_{nc}}))$ is a block rotation matrix contributing nothing to equation (1). Thus, it is possible that the estimated rotation matrix is not an element of *SO*(3). To demonstrate this, we test different values of Δ_1 by fixing Δ_2=0.01 and *sp*=100. The Frobenius norm between the estimated rotation matrix and its closest element in *SO*(3) is computed for each image denoted as R_Δ, then the logarithm for the largest R_Δ is computed, the result is shown in Fig. 4. As can be seen, the estimated rotation matrix tend to be further away from *SO*(3) as Δ_1 increases.

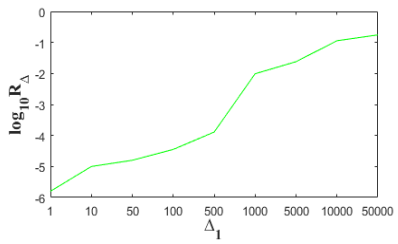


Figure 4. Block rotation constraint of various settings on Δ_1.

Based on this evaluation and to obtain a reliable and accurate solution for equation (1), we set Δ_1 = 100, Δ_2 = 0.01 and *sp*=100 in our all experiments.

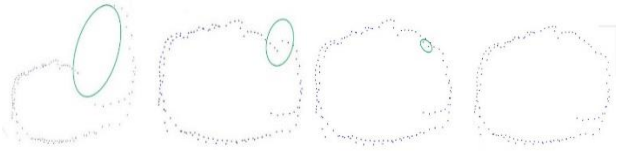
5.2 Experiments on terrestrial close range datasets

5.2.1 *Building* dataset

Our hybrid method classifies the triplets of the OMCTS into collinear and non-collinear ones and processes them separately. To show that this strategy is superior to the idea of considering all detected triplets as either non-collinear or collinear, we conduct experiments on the *building* dataset using three corresponding pipelines: hybrid, all non-collinear and all collinear (they are indicated by "HM", "ANC" and "AC", respectively, henceforth). As this dataset does not have ground truth exterior orientation and Wang et al. (2019b) was demonstrated to provide a reliable result for it, we use the exterior orientation from Wang et al. (2019b) as reference.

[1] https://github.com/wx7531774.

[2] https://3dom.fbk.eu/3domcity-benchmark.



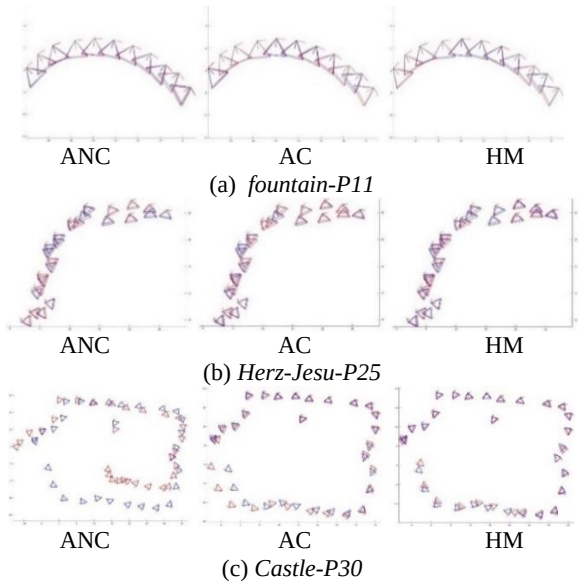
AN AC HM Reference

Figure 5. Motion trajectories of *Building* of different pipelines.

Fig. 5 shows the set of projection centres using the different pipelines (without refinement of bundle adjustment). The green ellipses denote drifts; the larger the ellipse, the bigger the drift, this in other words implies that "ANC" produces the worse result. The reason is that some triplets of the dataset are (nearly) collinear, which violates the non-collinear constraint described in corollary 1, thus, the corresponding estimated 3-Image essential matrix is not scale consistent. "AC" performs better than "ANC". We find that the method described in section 4.3 can actually also be used for non-collinear triplets. However, errors stemming from inaccurate Z values of object points in equation (10) can accumulate in the process of traversing all connected triplets as described in section 4.4, and this can lead to the drift depicted in Fig. 5. "HM" generates the best result, the detected non-collinear triplets satisfy the non-collinearity constraint of "ANC" and the remaining collinear triplets (less triplets compared to "AC") show less error accumulation. In addition, the method for solving collinear triplets only use two necessary relative orientations, which is not be as robust as solving non-collinear triplets using all the relative orientations. Thus, among these three pipelines, based on the result presented our hybrid method is the best one to deal with datasets consisted of both non-collinear and collinear images.

5.2.2 Three benchmark datasets with ground truth

We also inspected three benchmark datasets with ground truth of exterior pose parameters, namely, *fountain-P11*, *Herz-Jesu-P25* and *castle-P30* (Strecha et al., 2008). The interior orientation parameters are extracted from the EXIF information. Similar to the *Building* dataset, we ran the three pipelines ("ANC", "AC" and "HM") for these three benchmarks. Besides, we further compared our results to the results of several recent global rotation and translation estimation methods.



ANC AC HM

(a) *fountain-P11*

ANC AC HM

(b) *Herz-Jesu-P25*

ANC AC HM

(c) *Castle-P30*

Figure 6. Motion trajectory of three benchmarks with different pipelines, red triangles denote the results computed from corresponding pipelines and blue triangles indicate the ground truth exterior parameters.

Fig. 6 shows the results for the exterior orientation parameters of these three benchmarks by using the corresponding different pipelines (without bundle adjustment), where the blue triangles represent ground truth and the red triangles indicate the estimated exterior pose parameters (the estimated exterior pose parameters are transformed into the coordinate system of ground truth using the 3D similarity transformation method presented in Wang et al. (2019b)). From Fig. 6, we find that all three pipelines work very well on *fountain-P11*and *Herz-Jesu-P25*, as the blue and red triangles are very close to each other and some almost overlap. However, results of *castle-P30* look different, a similar phenomenon as described above for the building benchmark: "AC" is better than "ANC", and the proposed method "HM" is the best. This can be explained by the fact that the images of *fountain-P11*and *Herz-Jesu-P25* are all almost non-collinear and the relative orientations are already rather accurate, so error accumulation is not a major problem, thus, all three pipelines perform very well. However, *Castle-P30* is closer to *building* in that it has both collinear and non-collinear triplets, and outliers of relative orientations exist due to repetitive structures.

Visualizations of the results in Fig. 5 and Fig. 6 are provided for a qualitative comparison of the different pipelines. To generate a numerical analysis, based on the three benchmarks with ground truth we calculate the mean rotation error denoted as mean angle error and the mean translation error which are both listed in Tab. 1. From this table, it can be inferred that the exterior orientation parameters (rotation and translation) estimated by "ANC", "AC" and "HM" achieve nearly the same accuracy on *fountain-P11* and *Herz-Jesu-P25*, respectively. The result of *castle-P30* shows a very explicit superiority of "HM": the angle and translation error of our hybrid method are approximately 15 to 20 and 5 to 10 times smaller than those of "ANC" and "AC", respectively. What Tab. 1 implies is consistent with Fig. 6, thus, we can conclude that both "ANC" and "AC" can perform very well on small datasets with very few collinear triplets such as *fountain-P11*and *Herz-Jesu-P25* (as Fig. 6 (a) and (b) illustrate), whereas, for the *castle-P30* dataset with not only more images but also both, collinear and non-collinear images (see Fig.6 (c)), "ANC" results are invalid due to the non-collinearity constraint requirement, and the performance of "AC" also decreases because error accumulation increases, when more connected triplets are traversed. As in the first test, "HM" provides the best solution for the problem at hand. A visualization of image orientation and sparse 3D object point result after bundle adjustment is shown in Fig. 7.

Figure 7. Visualization of benchmarks' SfM results by "HM" (after bundle adjustment). Colourful triangles denote exterior pose parameters; red dots are estimated 3D object points.

To obtain a deeper understanding of the performance of "HM", we compare rotation and translation results of "HM" with those of several global rotation estimation and global translation estimation methods, respectively. Tab. 2 presents numerical results for the mean rotation and translation errors of different methods. Before bundle adjustment, "HM" outperforms all the other methods listed in Tab.2, specifically, the mean angle errors and mean translation error of "HM" are the smallest on all these three benchmark datasets (except for the translation error of *castle-P30*, where Wang et al. (2019b) is 2 millimetres better than "HM" which is negligible). This is probably a consequence of the

fact that we only use some optimal triplets in the extracted OMCTS, the selection acts as a kind of blunder detection method for the relative orientations, whereas the other methods typically employ more redundant relative orientations, and may thus be negatively influenced by relative orientations not spotted as outliers. After bundle adjustment, both rotation and translation accuracies are improved on all benchmark datasets, and remaining differences are negligible.

	fountain-P11			*Herz-Jesu-P25*			*castle-P30*		
	ANC	AC	HM	ANC	AC	HM	ANC	AC	HM
R	0.161	0.159	**0.156**	**0.186**	0.189	0.191	5.732	4.643	**0.277**
T	0.020	0.022	**0.019**	0.033	0.033	**0.028**	3.794	1.573	**0.155**

Table 1. Mean angle error R in degree and mean translation error T in meter for different pipelines. We highlight the best results of each dataset.

R	before bundle adjustment					after bundle adjustment			
	HM	Global_R	GR_L2	(1)	(2)	HM	Global	G_L2	
fountain-P11	**0.156**	0.251	0.261	0.249	0.45	**0.042**	0.136	0.140	
Herz-Jesu-P25	**0.191**	0.238	0.365	0.206	0.39	**0.023**	0.053	0.048	
castle-P30	**0.277**	0.745	0.954	0.583	0.96	**0.084**	0.133	0.129	
T	**before bundle adjustment**						**after bundle adjustment**		
	HM	Global_T	GT_L2	(1)	(2)	(3)	HM	Global	G_L2	(3)
fountain-P11	**0.019**	0.035	0.041	0.035	0.072	0.037	**0.008**	0.010	0.010	0.011
Herz-Jesu-P25	**0.028**	0.085	0.131	0.083	0.061	0.077	**0.013**	0.014	**0.013**	0.015
castle-P30	0.155	0.161	0.194	1.312	1.620	**0.153**	**0.019**	**0.019**	0.020	0.022

Table 2. Mean angle error R in degree and mean translation error T in meter for different global estimation methods. We compared our rotation results with Chatterjee and Govindu (2013) (Global_R), Reich and Heipke (2016) (1) and Jiang et al. (2015) (2). GR_L2 adopts the "Global_R" method with L2 norm, their corresponding results are provided by Wang et al. (2018). The translation results are compared with Reich and Heipke (2016) (1), Jiang et al. (2015) (2), Wang et al. (2019b) (3) and Wang et al. (2019a) using L1 and L2 norm denoted as Global_T and GT_L2, respectively. Note that the results of (1), (2) and Wang et al. (2019a) are directly cited from the corresponding papers, and we reimplemented the approach of Wang et al. (2019b). The best results of each dataset are highlighted.

5.3 Experiments on oblique aerial image dataset

Figure. 8 Overall view of the simulated urban scenario.

	before bundle adjustment			after bundle adjustment		
	RMS(x)	RMS(y)	RMS	RMS(x)	RMS(y)	RMS
(I)	-	-	-	0.140	0.147	0.204
HM	**2.199**	**3.437**	**4.474**	**0.132**	**0.138**	**0.191**
(II)	2.234	3.617	4.591	**0.132**	**0.138**	**0.191**
(III)	2.354	3.444	4.778	**0.132**	**0.138**	**0.191**

Table. 3 Precision assessment. RMS(x), RMS(y) and RMS are the RMS (root mean square) of reprojection residuals (in pixels) in image space in horizontal direction, vertical direction and Euclidean residual.

	before bundle adjustment			after bundle adjustment		
	CH1	CH2	CH3	CH1	CH2	CH3
(I)	-	-	-	**-0.340**	**-1.046**	**0.333**
HM	9.434	**14.52**	**8.191**	-0.915	1.462	0.533
(II)	9.232	17.32	7.969	-0.841	1.482	0.606
(III)	**9.116**	15.87	8.492	-0.761	1.503	0.685

Table. 4 Accuracy assessment in 10^{-1} mm. CH1, CH2 and CH3 are the corresponding check bars showed in Fig. 8.

To further explore the capability of our method, we test another dataset of oblique quasi-aerial images (Özdemir et al., 2019). This dataset includes a set of 420 nadir and oblique images (6016×4016 pixels each) captured in a controlled environment over an ad-hoc 3D test field which simulates a typical urban scenario, as shown in Fig. 8. Three evaluation criteria are proposed to assess the image orientation results: 1. Precision assessment, the reprojection residuals of 115 targets (red crosses in Fig. 8) are used to evaluate the precision of orientation results in image space; 2. Accuracy assessment, three control bars (shown as blue lines in Fig. 8) and three check bars (showed as yellow lines in Fig. 8) with known length are provided to evaluate the accuracy of the orientation results; 3. Relative accuracy assessment, the errors of translation and rotation are evaluated by taking the provided exterior pose parameters as a reference. More information is provided by Özdemir et al. (2019).

	before bundle adjustment					
	RMSE (X)	RMSE (Y)	RMSE (Z)	RMSE (O)	RMSE (P)	RMSE (K)
HM	**13.59**	54.78	**16.56**	**10.65**	**10.19**	**10.84**
(II)	14.66	**53.58**	19.63	11.73	13.35	13.99
(III)	18.14	55.96	17.95	11.73	13.35	13.99
	after bundle adjustment					
	RMSE (X)	RMSE (Y)	RMSE (Z)	RMSE (O)	RMSE (P)	RMSE (K)
HM	**5.645**	24.833	**8.064**	**1.495**	1.444	**2.455**
(II)	5.665	**24.824**	8.334	1.496	**1.435**	2.461
(III)	5.687	25.413	8.237	1.520	1.440	2.457

Table. 5 Relative accuracy assessment. Taking the exterior pose parameters of Özdemir et al. (2019) as a reference, RMSE (X), (Y) and (Z) are the root mean square error of translation parameters which is in 10-1 mm, RMSE (O), (P) and (K) are the root mean square error of three rotation angles (O, P and K denote Omega, Phi, Kappa, respectively) which is in degrees.

The corresponding evaluation criteria are listed in Tab. 3, 4 and 5, where the results of Özdemir et al. (2019) are denoted by (I); for this method results prior to bundle adjustment do not exist), Wang et al. (2019a) using L1 norm is (II) and Wang et al. (2019b)

is indicated as (III), "-" means the corresponding items are not available. Before bundle adjustment, among these methods we find that "HM" typically generates the best results (as the highlighted items show in these three tables) with only small discrepancies; this finding is basically identical with the results of *castle-P30* shown in Tab. 2. "ANC" and "AC" were also tested, however, "ANC" failed when solving equation (1) due to the collinearity of projection centres, see. Fig. 9, and "AC" is also not reliable as we have explained in the last section. The rotation error of (II) and (III) are identical because both of them apply the same method of Chatterjee and Govindu (2013) to estimate global rotations. After bundle adjustment, "HM", (II) and (III) achieve nearly the same precision. (I) is not included in Tab. 5, because the exterior orientation parameters of (I) are the reference for the relative accuracy assessments. The results after bundle adjustment have been published in a contest of image orientation[3]. Fig. 9 is the visualization of image orientation and 3D object points using "HM" from two different perspectives.

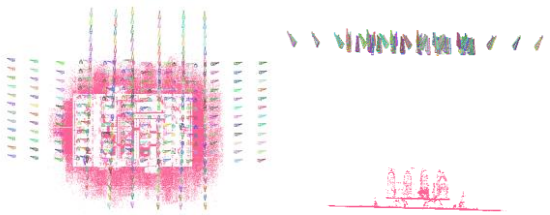


Figure. 9 Visualization of SfM results by using "HM".

6. CONCLUSIONS

In this paper, we present a novel hybrid global image orientation method which can solve global rotation and translation simultaneously. Specifically, an optimal minimum cover connected triplet set (OMCTS) is extracted, among which non-collinear and collinear triplets are first solved individually and global exterior pose parameters are then estimated by traversing all these solved connected triplets. Comparisons with several recent global SfM methods on different benchmarks demonstrate that our method can normally provide the best initial estimation of exterior orientation parameters for bundle adjustment. In the future, we will test larger and more interesting datasets, such as images downloaded from Internet (Wilson and Snavely, 2014), as these images are normally unordered which can create additional challenges for extracting the OMCTS. Also, the comparisons before applying final bundle adjustment and the time efficiency of the proposed hybrid method needs to be further investigated.

APPENDIX

1. Corollary 3. Given a non-collinear triplet, the corresponding scaled consistent 3-Image essential matrix is invariant to scales.

Proof. Assume E is a scale consistent 3-Image essential matrix, according to corollary 1 the block matrices of E can be denoted as $E_{ij} = \alpha_i R_i^T (T_i - T_j) R_j \alpha_j$. Then, let $\bar{E}$ be a 9×9 matrix whose corresponding block matrices are indicated as $\bar{E}_{ij} = \beta_{ij} E_{ij}$, where β_{ij} is a non-zero arbitrary positive scale factor. For a 3-Image essential matrix, we have three arbitrary scale factors β_{12}, β_{13}, β_{23}. Next, we show that for these three arbitrary scale factors it is possible to compute a new scalar for each image, whereas, it is not doable for an N-Image essential matrix with N>3.

[3] Find more details at http://3dom.fbk.eu/3domcity-task-1-result

1) N = 3, the goal is to obtain new scalars $\gamma_1, \gamma_2, \gamma_3$ s. t. they fulfill the new scale consistent 3-Image essential matrix $\bar{E}$

$$\begin{matrix}\gamma_1 \cdot \gamma_2 = \alpha_1\alpha_2\beta_{12} \\ \gamma_1 \cdot \gamma_3 = \alpha_1\alpha_3\beta_{13} \\ \gamma_2 \cdot \gamma_3 = \alpha_2\alpha_3\beta_{23}\end{matrix} \Rightarrow \begin{bmatrix}1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1\end{bmatrix}\begin{bmatrix}q_1 \\ q_2 \\ q_3\end{bmatrix} = \begin{bmatrix}\log\beta_{12} \\ \log\beta_{13} \\ \log\beta_{23}\end{bmatrix}$$ where $q_i = log\frac{\gamma_i}{\alpha_i}$ (i =1,2,3), $PM = \begin{bmatrix}1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1\end{bmatrix}$

In this case, we find that *rank* (*PM*) = 3 which implies we can compute q_i and get γ_i, and $\gamma_1 = \alpha_1\sqrt{\frac{\beta_{12}\beta_{13}}{\beta_{23}}}$, $\gamma_2 = \alpha_2\sqrt{\frac{\beta_{12}\beta_{23}}{\beta_{13}}}$, $\gamma_3 = \alpha_3\sqrt{\frac{\beta_{13}\beta_{23}}{\beta_{12}}}$.

2) N ≥ 4, we assume N = 4, similar to N =3. We can also set up equations similar to *PM* when N = 3 and the corresponding matrix $PM = \begin{bmatrix}1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1\end{bmatrix}$ and *rank* (*PM*) = 4, there is an infinite number of solutions and we cannot obtain a unique closed form solution like the case of N = 3, this is also true when N > 4.

Hence, the new 3-Image essential matrix $\bar{E}$ is also scale consistent, which means that a scale consistent 3-Image essential matrix is invariant to scalars.

2. Algorithm for generating the minimum cover connected image triplet set

Input Original exhaustive triplet set, each triplet's quality indicator, corresponding set of images $I_n = \{1, 2, 3, \ldots, n\}$.

Output Optimal minimum cover connected image triplet set

1. Build a triplet graph $G_\tau = \{\tau, \varepsilon_t\}$, where τ is the original exhaustive triplet set denoted as nodes and ε_t are the edges between triplets (two triplets are connected only if they share two common images).
2. Sort all triplets by their quality indicators in descending order, obtain corresponding triplet index set *Ind*.
3. Start with the triplet of largest quality indicators:
   Do {
   Remove τ_{Indj} and its corresponding edges from G_τ, then, check that:

   a. The remaining G_τ is connected;
   b. The images' number of remaining G_τ doesn't reduced.

   If both *a* and *b* fulfil, G_τ is successfully reduced by removing the corresponding triplet τ_{Indj}, otherwise, we keep G_τ unchanged and try the next iteration by considering $j=j+1$;
   }while (j = {1,2,3,…size of (*Ind*)})

Finally, the triplets which exist in the remaining G_τ consist of the triplet set that we desire.

3. Solving similarity transformation between two connect triplets

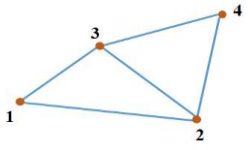


Figure 10. Two connected triplets.

Given two connected triplets shown by Fig. 10, image 2 and 3 are the shared images, $\{R_1, R_2, R_3\}$, $\{t_1, t_2, t_3\}$ and $\{R'_2, R'_3, R'_4\}$, $\{t'_2, t'_3, t'_4\}$ are the exterior pose parameters calculated from these two triplets, respectively. Our goal is to find a similarity transformation which can bring the second triplet to the coordinate system of first triplet.

For the relative rotation of the similarity transformation two solutions can be generated, i.e., $R_r = R_2\, R_2'^T$ or $R_r = R_3\, R_3'^T$. Then, we obtain a mean rotation matrix $\overline{R_r} = (R_2\, R_2'^T + R_3\, R_3'^T)/2$, and project $\overline{R_r}$ to the space of $SO(3)$ by SVD, i.e., $\overline{R_r} = U\Lambda V^T$, finally, $R_r = UI_{3\times3}V^T$. The scale factor λ_s and translation t_s of the similarity transformation are solved by

$$t_2 = \lambda_s R_r t'_2 + t_s,\ t_3 = \lambda_s R_r t'_3 + t_s \tag{12}$$

As each translation t_i has three entries, there are 6 equations and four unknowns in equation (12), least square is then used to obtain an optimal solution. Finally, image 4 can be transformed to be consistent with image 1,2 and 3 by $R_4 = R_r R'_4$, $t_4 = \lambda_s R_r t'_4 + t_s$.

ACKNOWLEDGEMENTS

We want to thank Christian Heipke for his invaluable recommendations. The author Xin Wang would like to thank the China Scholarship Council (CSC) for financially supporting his PhD study at Universität Hannover, Germany.

REFERENCES

Agarwal, S., Mierle, K. et al. (2007): Ceres Solver. http://ceres-solver.org (accessed 08.05.2017).

Agarwal, S., Snavely, N., Simon, I., Seitz, S. M., and Szeliski, R., 2009. Building Rome in a day. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp.72-79.

Arie-Nachimson, M., Kovalsky, S.Z., Kemelmacher-Shlizerman, I., Singer, A., Basri, R., 2012. Global motion estimation from point matches. In: Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), pp. 81–88.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine learning*, 3(1), pp. 1–122.

Bourmaud, G., Megret, R., Giremus, A. & Berthoumieu, Y., 2014. Global motion estimation from relative measuremnts using iterated extended Kalman filter on matrix Lie groups. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 3362–3366.

Chatterjee, A., Govindu, V.M., 2013. Efficient and robust large-scale rotation averaging. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 521–528.

Cui, Z., Tan, P., 2015. Global Structure-from-Motion by Similarity Averaging. Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 864-872.

Farenzena, M., Fusiello, A., and Gherardi, R., 2009. Structure-and-motion pipeline on a hierarchical cluster tree. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshop, pp. 1489-1496.

Govindu, V.M., 2001. Combining two-view constraints for motion estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2, pp. II–218.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press.

Hartley, R., Aftab, K., Trumpf, J., 2011. L1 rotation averaging using the Weiszfeld algorithm. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 3041–3048.

Jiang, N., Cui, Z., Tan, P., 2013. A global linear method for camera pose registration. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 481–488.

Jiang, N., Lin,W.-Y., Do, M. N. and Lu, J., 2015. Direct structure estimation for 3d reconstruction. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2655–2663.

Kasten, Y., Geifman, A., Galun, M., Basri, R., 2019a. GPSfM: Global Projective SFM Using Algebraic Constraints on Multi-View Fundamental Matrices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Kasten, Y., Geifman, A., Galun, M., Basri, R., 2019b. Algebraic Characterization of Essential Matrices and Their Averaging in Multiview Settings. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV).

Martinec, D., Pajdla, T., 2007. Robust rotation and translation estimation in multiview reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8.

Mayer, H., 2014. Efficient hierarchical triplet merging for camera pose estimation. In: German Conference on Pattern Recognition–GCPR 2014. Springer, Berlin, pp. 99–409.

Moulon, P., Monasse, P., Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV).

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (6), pp.756–770.

Özdemir, E., Toschi, I. and Remondino, F., 2019. A multi-purpose benchmark for photogrammetric urban 3D reconstruction in a controlled environment. ISPRS Archives of Photogrammetry Remote Sensing & Spatial Informa, Vol. XLII-1/W2, pp. 53–60.

Reich, M., Heipke, C., 2016. Convex image orientation from relative orientations. ISPRS Annals of Photogrammetry Remote Sensing & Spatial Informa, III-3, pp.107-114.

Reich, M., Yang, M. Y., Heipke, C., 2017. Global robust image rotation from combined weighted averaging. *ISPRS Journal of Photogrammetry & Remote Sensing*, 127, pp.89-101.

Schönberger, J. L., Frahm, J. M., 2016. Structure-from-Motion Revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Shah, R., Chari, V. & Narayanan, P. J., 2018. View-graph Selection Framework for SfM. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 553-568.

Snavely, N., Seitz, S. M., & Szeliski, R., 2006. Photo tourism: exploring photo collections in 3d. *Acm Transactions on Graphics*, 25(3), pp.835-846.

Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U., 2008. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8.

Toldo, R., Gherardi, R., Farenzena, M., & Fusiello, A., 2015. Hierarchical structure-and-motion recovery from uncalibrated images. *Computer Vision & Image Understanding*, 140, pp.127-143.

Wang, X., Rottensteiner, F., Heipke, C., 2018. Robust image orientation based on relative rotations and tie points. ISPRS Annals of Photogrammetry Remote Sensing & Spatial Informa, IV-2, pp. 295-302.

Wang, X., Rottensteiner, F., Heipke, C., 2019a. Robust Structure from Motion based on relative rotations and tie points. *PE&RS*, 5, pp.347-359.

Wang, X., Rottensteiner, F., Heipke, C., 2019b. Structure from Motion for ordered and unordered image sets based on random k-d forests and global pose estimation. *ISPRS Journal of Photogrammetry & Remote Sensing*, 147, pp. 19-41.

Wang, X., Xiao, T., Gruber, M., Heipke, C., 2019c. Robustifying relative orientations with respect to repetitive structures and very short baselines for global SfM. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW).

Wilson, K., Snavely, N., 2014. Robust global translations with 1DSFM. In: Proceedings of the European Conference on Computer Vision (ECCV). Springer, pp. 61–75.

Wu, C. 2013. Towards Linear-Time Incremental Structure from Motion. In: Proceedings of the IEEE Conference on 3dtv, pp.127-134.

Zach, C., Klopschitz, M., Pollefeys, M., 2010. Disambiguating visual relations using loop constraints. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1426–1433.