

Subtleties of extrinsic calibration of cameras with non-overlapping fields of view

Li, Zaijuan; Willert, Volker
(2019)

DOI (TUprints): <https://doi.org/10.25534/tuprints-00014532>
License: only the rights of use according to UrhG
Publication type: Article
Division: 18 Department of Electrical Engineering and Information Technology
Original source: <https://tuprints.ulb.tu-darmstadt.de/14532>

Zaijuan Li* and Volker Willert

Subtleties of extrinsic calibration of cameras with non-overlapping fields of view

Optimierungsdetails bei der extrinsischen Kalibrierung von Multi-Kamera-Systemen ohne überlappende Sichtfelder

<https://doi.org/10.1515/teme-2019-0030>

Received March 14, 2019; accepted May 6, 2019

Abstract: The calibration of the relative pose between rigidly connected cameras with non-overlapping fields of view (FOV) is a prerequisite for many applications. In this paper, the subtleties of the experimental realization of such calibration optimization methods like in (Z. Liu, et al., *Measurement Science and Technology*, 2011, Z. Li, V. Willert, *Intelligent Transportation Systems (ITSC)*, 2018) are presented. Two strategies that could be adapted to certain optimization processes to find better local minima are evaluated. The first strategy is a careful measurement acquisition of pose pairs for solving the calibration problem, which improves the accuracy of the initial value for the following non-linear refinement. The second strategy is the introduction of a quality measure for the image data used for the calibration, which is based on the projection size of the known planar calibration patterns on the image. We show that introducing an additional weighting to the optimization objective chosen as a function of that quality measure improves calibration accuracy and increases robustness against noise. The above strategies are integrated into different setups and their improvement is demonstrated both in simulation and real-world experiment.

Keywords: Extrinsic calibration, multi-camera system, quality measure.

Zusammenfassung: Die Kalibrierung der Relativpose zwischen starr verbundenen Kameras, die keine überlappenden Sichtfelder besitzen, ist eine notwendige Voraussetzung für viele Anwendungen der Bildverarbeitung. Der vorliegende Artikel bespricht die technischen Details, die bei der experimentellen Umsetzung der Kalibrieremethoden nach (Z. Liu, et al., *Measurement Science and Technology*, 2011, Z. Li, V. Willert, *Intelligent Transportation Sys-*

tems (ITSC), 2018) beachtet werden müssen, um genaue Kalibrierergebnisse zu erhalten. Es werden zwei Strategien vorgestellt, welche es dem Optimierungsprozess ermöglicht, bessere lokale Minima von nichtkonvexen Gütefunktionen, die zur Kalibrierung benutzt werden, zu finden. Die erste Strategie behandelt die Aufnahme und Auswahl von Messungen von geeigneten Bilderpaaren, wodurch bessere Initialwerte zur Lösung des nichtkonvexen Optimierungsproblems erzeugt werden können. Die zweite Strategie stellt ein Gütemaß auf Basis der Größe der reprojizierten Fläche des Kalibrierkörpers in den zur Kalibrierung verwendeten Bildaufnahmen vor. Dieses Maß kann als zusätzliche Gewichtung in der Gütefunktion verwendet werden und erzeugt genauere Kalibrierergebnisse, die robuster gegen Fehler auf Bildkoordinatenmessungen ausfallen. Beide Strategien werden für unterschiedliche Kamerakonfigurationen sowohl simulativ, als auch anhand echter Messdaten evaluiert.

Schlagwörter: Extrinsische Kamerakalibrierung, Multi-Kamera-System, Qualitätsmaß.

1 Introduction

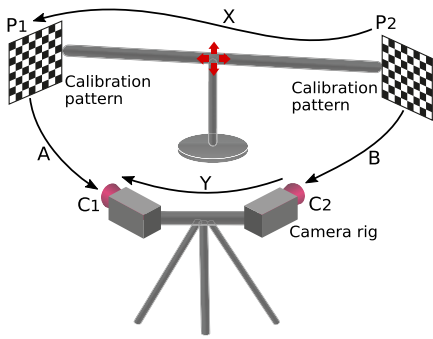
It is ubiquitous to have a multi-camera system in different research areas. For example, in the robot community: A multi-camera platform provides more flexibility in sensor placement on a mobile robot for simultaneous localization and mapping (SLAM) [3]; The majority of the quadcopters are equipped with at least two cameras to perform automatic localization and landing tasks [4]. In the car industry, the multi-camera infrastructure could be used for computing loop-closure constraints [5], or be integrated to assist effective parking [6]. In human-machine interaction area, a multi-camera system enables people to interact with the environment by building a real-time 3D modeling of it [7] or could be applied to people-tracking function for virtual reality television studios [8]. Each of such camera systems needs an extrinsic camera calibration, which comprises the estimation of the relative

*Corresponding author: Zaijuan Li, Technische Universität

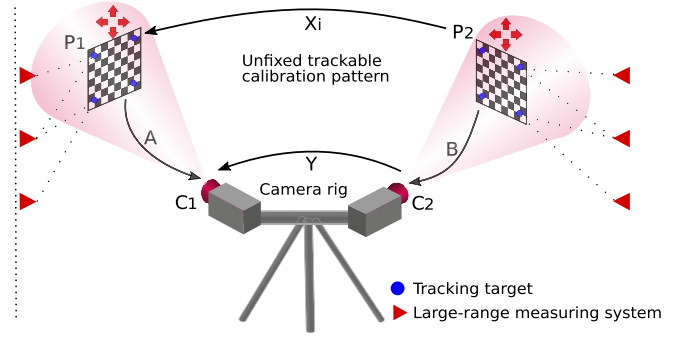
Darmstadt, Darmstadt, Germany, e-mail:

zaijuan.li@rnr.tu-darmstadt.de

Volker Willert, Technische Universität Darmstadt, Darmstadt, Germany



(a) Liu's setup [1].



(b) Large-range measuring device based setup [2].

Figure 1: Illustration of two setups that could apply the proposed strategies. The red arrows appearing in the above figures indicate that the objects with red arrows overlaid have to be moved or placed into different positions during the calibration procedure.

pose $\mathbf{X} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$ between cameras with non-overlapping FOV.

Different classifications of the existing calibration methods [1] [9], together with detailed analyses and discussion of the methods could be found in [2], [10]. In principle, the proposed strategies are suitable for any setup that builds its objective function based on the reprojection error of 3D-2D point correspondences constrained by 3D-3D closed-loop pose transformation $\mathbf{AX} = \mathbf{YB}$, where \mathbf{X} and \mathbf{Y} are the unknown transforms, and \mathbf{A} , \mathbf{B} are the estimated pose pair from measurements. In this work, the two setups shown in Figure 1 are taken as examples to demonstrate the integration of the proposed strategies.

Liu's setup [1] shown in Figure 1a uses a movable calibration device which rigidly links two planar calibration patterns $P1$ and $P2$ whose relative pose \mathbf{X} is unknown. The planar calibration pattern could either be a fiducial marker or a chessboard. By changing the pose of the calibration device relative to the camera rig, a set of images $\{\mathbf{I}_i^{P1}, \mathbf{I}_i^{P2}\}_{i=1}^n$ containing the calibration patterns is collected, based on which the relative pose pair $\{\mathbf{A}_i, \mathbf{B}_i\}_{i=1}^n$ between the camera pair $C1$, $C2$ and the corresponding calibration pattern $P1$, $P2$ could be recovered. The initial estimation of \mathbf{X} and \mathbf{Y} is calculated by solving $\mathbf{AX} = \mathbf{YB}$. In order to further improve the calibration accuracy, the initial value of \mathbf{X} and \mathbf{Y} is then applied to minimize the objective function, which is based on the sum of the reprojection errors ϵ_{ij} and ϵ_{il} , where i represents the measurement number, and j , l are the number of fiducial features used for estimating pose pair.

Similar to Liu's setup, the unfixed trackable pattern setup [2] demonstrated in Figure 1b introduces a highly accurate tracking system 'OptiTrack', which is used for accu-

rately localizing the two pattern boards after their coordinate frames being aligned with that of the tracking targets attached to them, and the pattern boards in this case are no longer rigidly linked and could be independently placed to the corresponding cameras. In this case, the 3D-3D closed-loop pose transformation is formulated as $\mathbf{AX}_i = \mathbf{YB}$, where \mathbf{X}_i is the i -th recovered relative pose between the calibration patterns using the tracking system. The extrinsics could thus be optimized using the reprojection error based objective.

Though Liu's setup needs extra infrastructure and additional interaction, the calibration patterns could be detected reliably with sub-pixel accuracy, which provides true scale information and could be further included in the optimization process. Meanwhile, the camera rig does not have to be moved during the calibration procedure, which is a big advantage, especially for mobile vehicles. However, the limited pose change space of the calibration targets could result in instability [10]. Applying a large-range measuring system is generally more accurate but the setup complexity and the costs are high.

2 Problem statement

The underlying instabilities, namely the **reduced, twisted** pose change space and the **imbalanced** measurement quality [10], are explained using Liu's setup, since this setup shows more prominent subtleties compared to the unfixed trackable pattern setup.

The measurement quality refers to the resolution of the captured calibration pattern. The measurement space or the pose change space is defined as a collection of pose

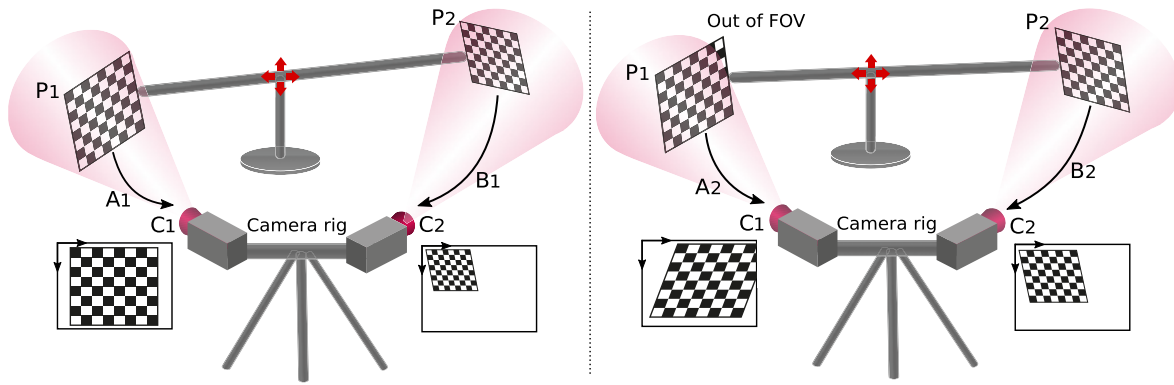


Figure 2: An example showing the twisted change in the resulting images after a minor pose adjustment. Since the measurement quality of the pattern $P2$ in the left figure is bad, the camera rig moves a little in the neighborhood of the current pose (A_1, B_1) in order to improve its quality. After the movement, even though the quality of the pattern $P2$ gets better, the calibration pattern $P1$ is no longer completely in the FOV of the camera $C1$.

pairs A_i and B_i . All the pose pairs in the measurement space meet the following conditions: 1) For each pose pair, both planar calibration items have to be in the FOV of the corresponding cameras such that all the coordinates of the projections could be extracted without outliers and with a certain accuracy; 2) The resulting measurement quality from each pose pair should be above a certain threshold.

One major practical issue to reach accurate calibration results is a proper set of accurately estimated measurement pairs $\{A_i, B_i\}_{i=1}^n$ covering all six degrees of freedom of the pose X and Y . However, collecting a proper set of measurement pairs in Liu's setup is problematic because of the rigid coupling between the two patterns, whose effect is disastrous. First and most straightforward, the closed-loop coupling reduces the pose change space since both of the calibration patterns have to appear within the FOV of the corresponding camera. The second consequence resulting from the coupling is the twisted pose change space. Due to the coupling $B_i = Y^{-1}A_iX$, a minor change in pose A_i would lead to a compound change in B_i . The same happens with a minor change in pose B_i . The twisted pose change space indicates the hardness of capturing both calibration patterns with high resolution (Figure 2). After a minor change in the pose of the calibration rig, the calibration pattern $P1$ is not completely covered by the FOV of the camera $C1$. Besides, this property also adds another layer of difficulty to the data collection process making it quite anti-intuitive: Though some pose pairs with good quality are theoretically valid, they are difficult to acquire in reality.

Another negative effect caused by the closed-loop constraint is the imbalanced measurement quality. Because of the coupling, the placement of one calibration pattern will influence the placement of the other one. Hence, acquiring a set of images that are of high resolution of both cali-

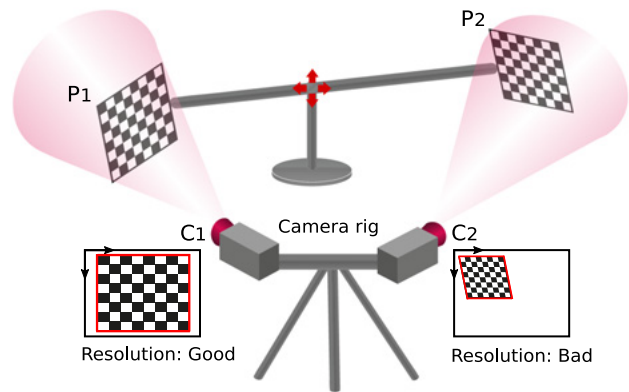


Figure 3: Relationship between the pose of the calibration rig relative to the camera pair and the corresponding resolution quality: When the calibration pattern from one side of the rig is placed near to the camera, an image with high resolution will be captured, while the calibration pattern from the other side would be captured with comparatively lower resolution and vice versa. Corresponding images containing different fiducial patterns are captured at the same measurement time, which are of different projection sizes. The projection size is defined as the area surrounded by the red lines on each image. In this example, the pattern $P1$ generates a larger projection size on the image, hence higher resolution than $P2$. When both images are corrupted by the same level of noise, the pose estimation using the measurement of $P1$ will be less sensitive to noise and will produce a better pose estimation.

bration patterns is challenging. Figure 3 demonstrates the relationship between the captured measurement quality and the relative pose between the calibration rig and the camera pair. Meanwhile, example images are presented which give imbalanced projection size of different calibration patterns for the same pose pair.

From the perspective of robustness, images containing the calibration patterns should be taken from as many dif-

ferent poses as possible. For accuracy consideration, they should be captured with as much resolution as possible since high resolution of the pattern results in better pose estimation. Generally, during the collection of the measurements, balances have to be kept between the pose pair variety and the measurement quality.

3 Data selection strategy and the weighted optimization method

3.1 Data selection strategy

The reprojection error based objective function needs an initial value of \mathbf{X} and \mathbf{Y} , which is estimated from solving $\mathbf{AX} = \mathbf{YB}$. The methods of solving the equation demand a set of pose pairs $\{\mathbf{A}_i, \mathbf{B}_i\}_{i=1}^n$ covering the six degrees of freedom.

As explained before, the pose change space in Liu's setup is reduced due to the closed-loop coupling, which would decrease the accuracy of solving $\mathbf{AX} = \mathbf{YB}$ if the spatial distribution of the collected pose pairs is not scattered enough. Besides, a minor change in the pose pair results in a compound, twisted change in the resulting images, which makes it less straightforward to distinguish the spatial difference between different pose pairs: Pose pairs that are spatially close might result in very different images. Thus, collected pose pairs need to be carefully handled, otherwise, they would bring potential hazard to the calibration stability.

On the other hand, the accuracy of the pose pair is also crucial to the solution of $\mathbf{AX} = \mathbf{YB}$: Better estimated pose pairs enhance the estimation results. Therefore, the measurement quality should work in conjunction with the measurement space in order to generate an optimal initial value of \mathbf{X} and \mathbf{Y} .

Normally, a large number of pose pairs are suggested during the data collection process with the purpose of covering as much measurement space as possible. An extra data selection filter is then applied to all the collected pose pairs. The filter calculates the rotational difference e^R and translational difference e^T between all the pose pairs based on the criteria formulated in (5), (6), as well as the projection size S^P of the calibration pattern. The pose pair $(\mathbf{A}_i, \mathbf{B}_i)$ whose rotational difference or translational differences are below certain thresholds θ^R, γ^T compared to all the rest pose pairs $(\mathbf{A}_j, \mathbf{B}_j) (j \neq i)$ will be excluded: $(e^R(\mathbf{A}_i, \mathbf{A}_j) < \theta^R \parallel e^R(\mathbf{B}_i, \mathbf{B}_j) < \theta^R \parallel e^T(\mathbf{A}_i, \mathbf{A}_j) < \gamma^T \parallel e^T(\mathbf{B}_i, \mathbf{B}_j) < \gamma^T)$. Besides, the pose pairs whose resulting projection size of all the captured calibration pattern

is smaller than some pre-defined value τ will not be included: $(S_i^{P1} < \tau \parallel S_i^{P2} < \tau)$.

In the end, a subset of the collected pose pairs which is chosen with more discretion is used to solve $\mathbf{AX} = \mathbf{YB}$, which provides a better estimated initial value of \mathbf{X} and \mathbf{Y} for the following non-linear optimization.

3.2 Weighted non-linear optimization method

The optimization problem of Liu's setup is formulated as follows:

$$(\hat{\mathbf{R}}_X, \hat{\mathbf{t}}_X, \hat{\mathbf{R}}_Y, \hat{\mathbf{t}}_Y) = \arg \min_{\mathbf{R}_X, \mathbf{t}_X, \mathbf{R}_Y, \mathbf{t}_Y} \left(\sum_{i=1}^n \sum_{j=1}^m \|\epsilon_{ij}^{P1}\|_2^2 + \sum_{l=1}^o \|\epsilon_{il}^{P2}\|_2^2 \right), \quad (1)$$

where $\mathbf{R}_X, \mathbf{t}_X, \mathbf{R}_Y, \mathbf{t}_Y$ are the rotational and translational matrices of the unknown relative pose \mathbf{X} and \mathbf{Y} . m, o stand for the number of the fiducial features from the corresponding patterns, and n is the number of the collected pose pairs. ϵ_{ij}^{P1} and ϵ_{il}^{P2} are the reprojection errors from different calibration patterns. This optimization problem is non-convex, so the iterative optimization can only guarantee to converge to a local minimum and a proper initialization is needed in order to reach a good estimation. In this work, the initial value of \mathbf{X} and \mathbf{Y} is calculated applying the method in [11] beforehand.

The underlying measurement imbalance discourages the objective function from including all the measurements and treating them equally. Considering the unpleasant imbalance, additional weightings λ_i^{P1} and λ_i^{P2} are introduced to the objective (1), which leads to:

$$(\hat{\mathbf{R}}_X, \hat{\mathbf{t}}_X, \hat{\mathbf{R}}_Y, \hat{\mathbf{t}}_Y) = \arg \min_{\mathbf{R}_X, \mathbf{t}_X, \mathbf{R}_Y, \mathbf{t}_Y} \left(\sum_{i=1}^n (\lambda_i^{P1} \sum_{j=1}^m \|\epsilon_{ij}^{P1}\|_2^2 + \lambda_i^{P2} \sum_{l=1}^o \|\epsilon_{il}^{P2}\|_2^2) \right). \quad (2)$$

The weighting λ_i^{P1} used for the reprojection error related to the pattern $P1$ is chosen to be the square root of the projection size S_i^{P2} of the pattern $P2$ normalized by the full image size S_{\max} :

$$\lambda_i^{P1} = \sqrt{S_i^{P2}/S_{\max}}. \quad (3)$$

The other weighting factor λ_i^{P2} is calculated in a similar way:

$$\lambda_i^{P2} = \sqrt{S_i^{P1}/S_{\max}}. \quad (4)$$

The reason for choosing such weighting lies in the replacement of \mathbf{A}_i with $\mathbf{YB}_i\mathbf{X}^{-1}$. The reprojection error produced from the calibration pattern $P1$ now depends on its

replacement $\mathbf{Y}\mathbf{B}_i\mathbf{X}^{-1}$ which has the pose estimation \mathbf{B}_i inside, so the estimation accuracy of \mathbf{B}_i influences the reprojection error of $PI: \mathbf{B}_i$ with better estimation quality should have more influence on the optimization results and this leads to a higher weight λ_i^{P1} . In this case, the projection size is regarded as an indirect indicator of the measurement quality. Larger projection size indicates better measurement quality. The same happens with the replacement of \mathbf{B}_i .

The integration of the method improves calibration accuracy and stability: The introduced weighting factor allows more pose pairs to be safely included in the calibration procedure since their influence on the estimation is now correlated with the quality indicator, namely the projection size of the captured calibration pattern.

4 Validation on simulated dataset

In this section, the data selection strategy and the weighted non-linear optimization method are validated on a synthetic dataset. First, the explanation of how synthetic data is generated for Liu's setup is provided, then the definition of error metrics which are going to be used for the evaluation of different algorithms is presented. In the end, state-of-the-art methods with different settings are implemented and compared. The method of applying the highly accurate tracking system is not implemented in the simulation since there is no appropriate noise model for the tracking system.

4.1 Synthetic dataset

As illustrated in Figure 1a, a customized calibration device is introduced to assist the calibration procedure, except that all the true transforms are exactly known in the simulation.

First, an exhaustive searching program is first run based on all the known ground truth such as the relative pose between the calibration patterns \mathbf{X} , the relative pose between the cameras \mathbf{Y} , the camera intrinsic parameters, etc. to produce a pose pair bank which consists of over 1,400 pose pairs. All pose pairs in the bank meet the following requirements: Each pose pair in the bank is different from the rest both in translation and rotation so that the pose pairs in the bank densely sample the whole measurement space, which is continuous; The projection size of the calibration pattern must exceed a certain threshold, which guarantees the minimum quality of the measurement. In this experiment, the threshold is set to 0.14

of the full image plane. The synthetic measurements are then generated based on the pose pair bank: The true pose pairs are first randomly extracted from the bank; The noise-free 2D coordinates obtained through the projection process are corrupted with Gaussian noise afterward; In the end, the noise-corrupted 2D coordinates are used as the measurement to recover the noisy pose pairs \mathbf{A}_i and \mathbf{B}_i .

To demonstrate the influence of the spatial distribution of pose pairs on the calibration results, measurement sets with the following characteristics could be generated from the pose pair bank: a). Spatially scattered pose pair set with larger projection size. b). Spatially clustered pose pair set with larger projection size. c). Scattered distributed pose pair set with smaller projection size. d). Clustered distributed pose pair set with smaller projection size. The difference between the scattered distribution and the clustered one is that the former has both larger rotation and translation differences among all the pose pairs. Because all generated pose pairs are extracted from the bank, each pose pair in the clustered set has at least the same minimum translational and rotational difference as the ones in the bank. Same for the measurement quality: The pose pairs that have larger projection size are extracted from the pose pair bank based on a larger projection size threshold. Though the quality of the pose pair set with smaller projection size is bad compared to the pose pair set with large projection size, the former set is still guaranteed the minimum required quality since they are generated from the bank. The combination of the pose pair distribution and the projection size gives four extreme measurement sets, namely scattered large, clustered large, scattered small, and clustered small, where clustered and scattered suggest the pose distribution while large and small means the projection size of the resulting calibration pattern. The codes for the calibration model as well as the optimization strategies are available online.¹

4.2 Error metric

$\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ represent the estimated solutions which are calculated by applying different calibration methods. The ground truth of \mathbf{X} and \mathbf{Y} is known in the simulation environment, so the estimated $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ could be directly compared based on the following error metrics. Since the error metric calculation of \mathbf{X} and \mathbf{Y} is the same, only \mathbf{X} is taken as the example.

¹ <https://github.com/zaijuan/eye-to-eye-calibration.git>

4.2.1 Rotation error

The method of Wunsch et al. in [12] is applied to define the rotation error. $\hat{\mathbf{q}}_X$ denotes the estimated quaternion of \mathbf{X} and \mathbf{q}_X the ground truth quaternion. The rotation error e_X^R is defined as:

$$e_X^R = \min\{\arccos(\mathbf{q}_X \cdot \hat{\mathbf{q}}_X), \pi - \arccos(\mathbf{q}_X \cdot \hat{\mathbf{q}}_X)\}, \quad (5)$$

in which ‘ \cdot ’ denotes the inner product of two quaternion vectors to be compared. Here the rotation error is represented by the angles returned by *arccos* and then mapped to $[0, 90^\circ]$.

4.2.2 Translation error

The estimated translation vector is described as $\hat{\mathbf{t}}_X$, and the ground truth is \mathbf{t}_X . The translation error is computed as follows:

$$e_X^t = \|\mathbf{t}_X - \hat{\mathbf{t}}_X\|. \quad (6)$$

4.3 Simulation results

In this part, the calibration results of different methods with different settings are presented. The calibration results of the method in [11], which is referred to as Wang’s method, will be presented since all the other non-linear methods take its estimation of \mathbf{X} and \mathbf{Y} as the initial value. Since different camera configurations will lead to different scales of the resulting calibration error, the ground truth of \mathbf{Y} is given as follows for better and clearer demonstration:

$$\mathbf{Y} = \begin{bmatrix} 0.5 & 0 & 0.866 & 1.04 \\ 0 & 1 & 0 & 0 \\ -0.866 & 0 & 0.5 & -0.6 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

To prove that applying the weighting factor during the optimization process alleviates the imbalance of the measurement quality, both non-weighted and weighted methods are implemented. The non-weighted method doesn’t utilize the weighting factor, while the weighted method applies the weighting factor. In parallel, the method of minimizing the reprojection error from only one calibration pattern is implemented for two reasons: To test whether it gives better estimation results than Wang’s method, which doesn’t minimize the reprojection error; By comparing to Liu’s method, the necessity of minimizing the reprojection error from both calibration patterns is validated.

The above methods are referred to as unweighted two-side constrained method (Liu), weighted two-side constrained method (Wgt-Liu), unweighted one-side constrained method (Unwgt-1), weighted one-side constrained method (Wgt-1).

Besides different methods, it’s also significant to show how their calibration robustness and accuracy correspond to the increase of image noise and various measurement numbers: In the first setting, the number of pose pairs changes from 5 to 45 with fixed Gaussian noise of 1.0 pixel; In the second one, the added Gaussian noise on the image varies from 0.2 to 1.4 pixels with a fixed number of 25 measurement pairs. The results shown below are the average taken of 100 iteration runs. For each iteration, the pose pairs are randomly extracted from the pose pair bank and processed applying different methods. Since the generation of each measurement set is random, 100 different measurement sets will be generated and used for the calibration procedure after repeating 100 times. Therefore, the demonstrated calibration results are the average of the overall calibration error instead of a specific measurement set, which is objective and reliable.

The estimation error of \mathbf{Y} under different settings is demonstrated in Figure 4 and Figure 5. Since the results of \mathbf{X} are similar to \mathbf{Y} in magnitude and pattern, it is unnecessary to present repetitive work.

The above experiment results validate the accuracy and robustness of the weighting factor strategy: The integration of the weighting factor generates better results regardless of different settings; With the increase of noise level, the benefits from applying the weighting factor become more noticeable.

It has been validated in the simulation that the final estimation has been improved by choosing a subset, whose pose pairs are comparatively scattered and measurements are of good quality. These results are not shown in the above figures because the improvement is not noticeable and would cover the results from Wgt-Liu. Instead, four extreme types of measurement sets, namely scattered good quality, clustered good quality, scattered bad quality, and clustered bad quality, are utilized to emphasize these differences depending on their spatial distribution and measurement quality. For each configuration, the measurement number is set to 25, and the noise level is 1.0 pixel.

For all methods, spatially scattered pose pairs with better measurement quality (larger projection size) generate the most accurate estimation, while clustered spatially pose pairs with smaller projection size lead to the worst estimation results. Scattered pose pairs with smaller projection size produce better results than clustered pose pairs

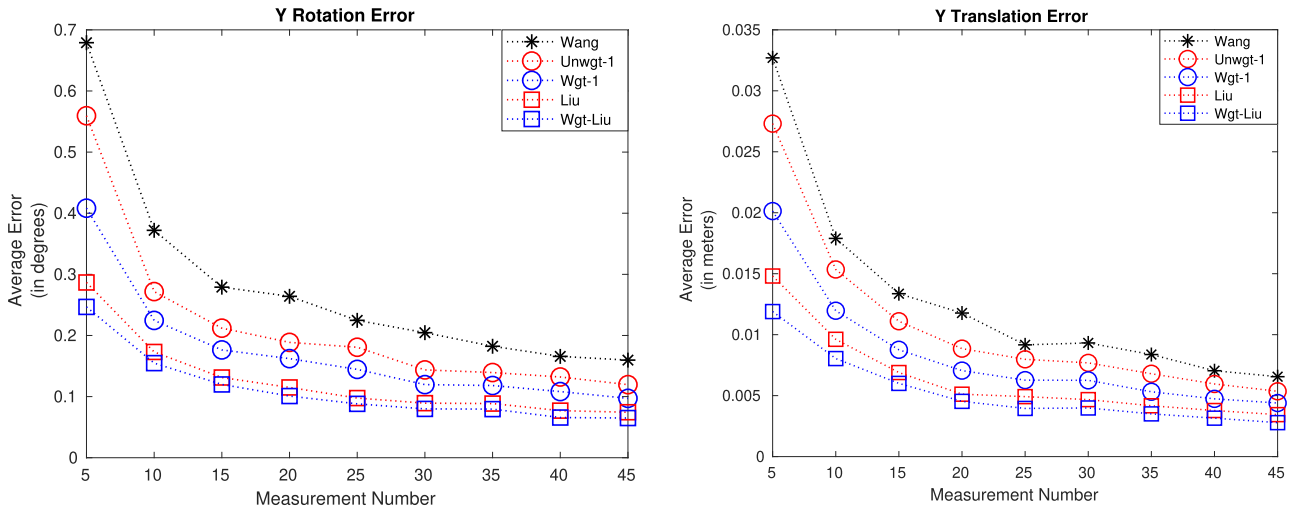


Figure 4: Estimation error of different methods with increased number of measurements.

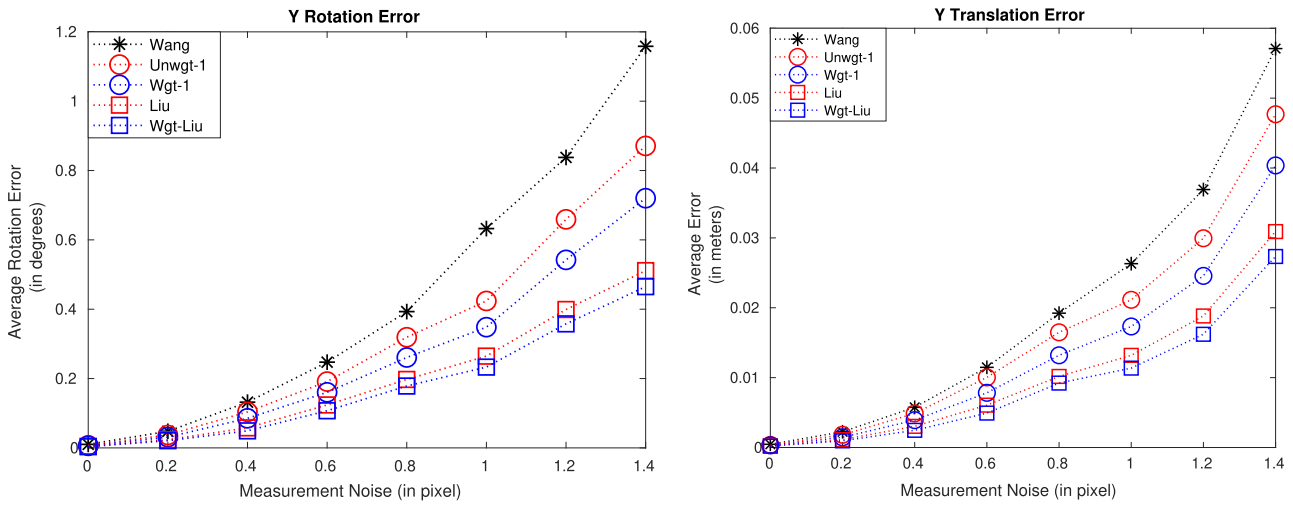


Figure 5: Estimation error with regard to increased image noise and different methods.

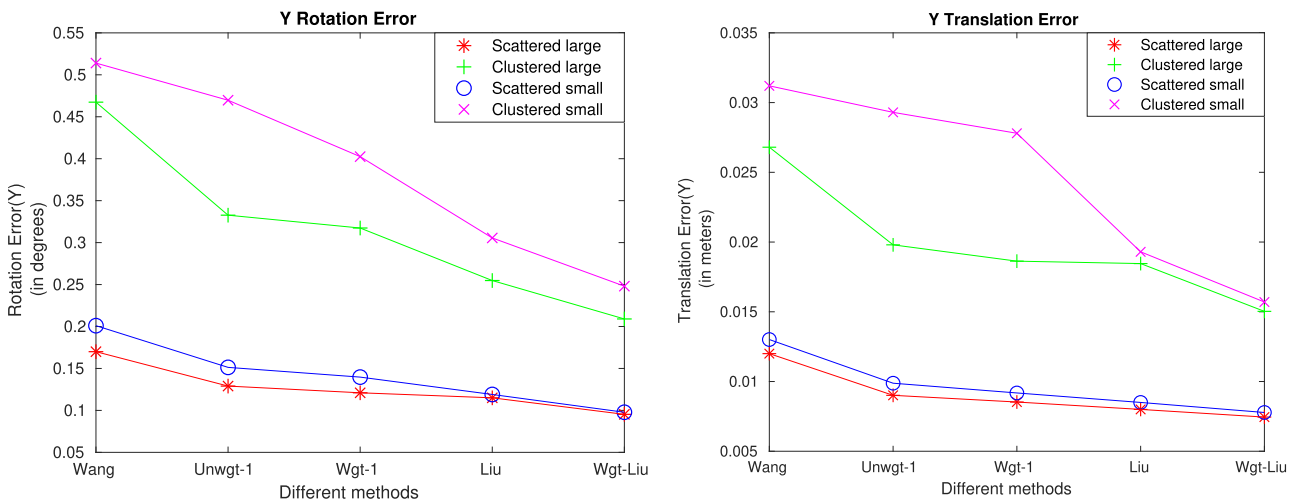


Figure 6: Estimation error of different methods with different pose pair configurations.

with larger projection size regardless of different methods, which implies that the calibration methods are more demanding on the distribution of pose pairs than their measurement quality.

The results bring some insights into the tradeoff between the spatial distribution of pose pairs and their generated measurement quality. It is crystal clear that the combination of scattered pose pair distribution and larger projection size produces the best calibration results. However, these two factors are to some extent mutually restricted: Scattered pose pair distribution implies the diversity of the projection size; While the demand of larger projection size limits the spatial distribution of pose pairs. This further explains why the introduced weighting factor is important during the optimization process: First, it increases the pose change space by allowing larger varying range of the measurement quality; Second, the increased measurement space helps to provide a more accurate initial value from solving $\mathbf{AX} = \mathbf{YB}$, which is used for the following weighted non-linear refinement.

5 Real experimental results

5.1 Experiment setup

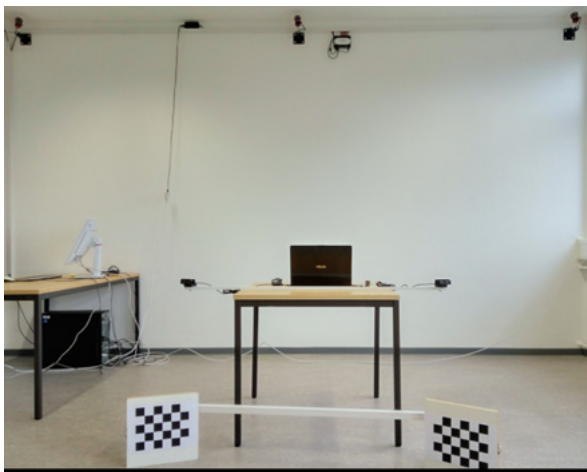
In the real experiment, Liu's setup [1], the fixed trackable pattern setup, as well as the unfixed trackable pattern setup [2] are implemented. The fixed trackable pattern

setup is a mixture of Liu's setup and the unfixed trackable pattern setup: The calibration patterns are rigidly linked as in Liu's setup, while the unknown relative pose \mathbf{X} between the calibration patterns could be recovered from the introduced highly accurate tracking system and directly used for the calibration procedure as in the unfixed trackable pattern setup.

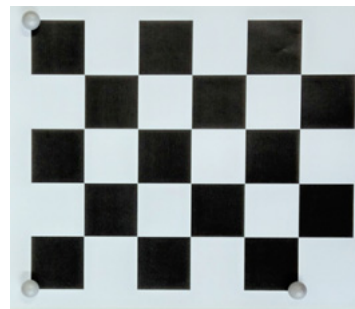
When a highly accurate tracking system is available in the calibration environment, the unfixed trackable pattern setup is preferred over the fixed trackable pattern setup since the former generates measurements with better quality than the latter one. By including the fixed trackable pattern setup in the real experiment, the subtleties between different calibration setups and optimization methods could be better revealed.

Figure 7a shows the real experiment environment, where a camera rig mounted with two cameras with non-overlapping FOV, an external calibration device, and an equipped highly accurate tracking system 'OptiTrack' are provided. The introduced calibration pattern boards used for recovering the relative pose to the camera pair could be accurately localized within the tracking system after aligning their coordinate frames with that of the tracking targets attached to them (Figure 7b).

The experiments are carried out as follows. Instead of hesitating which pose pair to include, it is preferable to collect an abundant amount of pose pairs covering as much measurement space as possible. In the end, a set of pictures $\{\mathbf{I}_i^{P1}, \mathbf{I}_i^{P2}\}_{i=1}^n$ containing the planar calibration pattern used for recovering $\{\mathbf{A}_i, \mathbf{B}_i\}_{i=1}^n$, and the recovered \mathbf{X} in the



(a) The real experiment environment.



(b) The tracking targets.

Figure 7: The left figure demonstrates the real experiment environment. Above is the equipped high accuracy tracking system 'OptiTrack'. The camera pair with non-overlapping FOV is rigidly connected and fixed in the experiment, and the calibration rig with two known planar patterns rigidly linked is placed on the ground. The right figure shows the calibration board used for detection by both the camera and the tracking system. The coordinate frame of the tracking targets and the pattern board coordinate system are aligned.

fixed trackable pattern setup or \mathbf{X}_i in the unfixed trackable pattern setup are collected. All the collected pose pairs are first filtered applying the data selection strategy and a subset is then used to different optimization processes.

5.2 Experimental results

Unlike in the simulation, there is no ground truth in the real experiment. In order to evaluate the calibration results from different setups and verify the improvement brought by the proposed optimization strategies, the unfixed trackable pattern setup with the integration of the weighting optimization method serves as the benchmark since this configuration generates the best possible calibration results.

The same error criteria are used as in the simulation to evaluate the calibration difference of different methods. The benchmark is set as the weighted estimation of the unfixed trackable pattern configuration. The term difference is used in the real experiment instead of error to indicate that although the ground truth of \mathbf{Y} is unknown, it could be estimated with the highest accuracy applying the unfixed trackable pattern configuration.

Figure 8 shows the calibration differences of different setups and different methods. Liu's method with the integrated optimization method generates a little larger but bearable calibration differences compared to the fixed trackable pattern configuration. Wang's method which doesn't minimize the reprojection error deviates the farthest from the benchmark. The performance of the fixed trackable pattern setup lies between Liu's setup and the

unfixed trackable pattern setup. In the unfixed trackable pattern setup, the difference between weighted and unweighted estimation is minor since in this case, all calibration patterns could be captured with relatively high resolution. Nevertheless, applying the weighting factor generates less different results compared to the benchmark regardless of different setups and methods.

6 Conclusion and future work

In this work, a weighted non-linear optimization method together with the data selection strategy which are applicable to certain calibration setups are developed. The optimization method introduces an extra quality measure factor to the objective function, which increases the measurement space and improves the calibration accuracy. Hence, instability could be alleviated and robustness could be safely guaranteed. Besides, by carefully choosing a measurement subset, the possibility of getting trapped in a worse local minimum is reduced.

During the simulation and the real experiment, the appropriate size of the calibration patterns and the corresponding relative pose \mathbf{X} between them are determined through trial and error. Since these two variables influence the measurement space and the measurement quality, they will also have an effect on the calibration results. In this paper, how to determine the optimal values of these two variables is not investigated.

References

1. Z. Liu, G. Zhang, Z. Wei, and J. Sun, "A global calibration method for multiple vision sensors based on multiple targets," *Measurement Science and Technology*, vol. 22, no. 12, p. 125102, 2011.
2. Z. Li and V. Willert, "Eye-to-eye calibration for cameras with disjoint fields of view (in press)," in *Intelligent Transportation Systems (ITSC)*. IEEE, 2018.
3. M. Kaess and F. Dellaert, "Probabilistic structure matching for visual slam with a multi-camera rig," *Computer Vision and Image Understanding*, vol. 114, no. 2, pp. 286–296, 2010.
4. E. Altuğ, J. P. Ostrowski, and C. J. Taylor, "Control of a quadrotor helicopter using dual camera visual feedback," *The International Journal of Robotics Research*, vol. 24, no. 5, pp. 329–341, 2005.
5. G. H. Lee, F. Fraundorfer, and M. Pollefeys, "Structureless pose-graph loop-closure with a multi-camera system on a self-driving car," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 564–571.
6. Y. Suzuki, M. Koyamaishi, T. Yendo, T. Fujii, and M. Tanimoto, "Parking assistance using multi-camera infrastructure," in

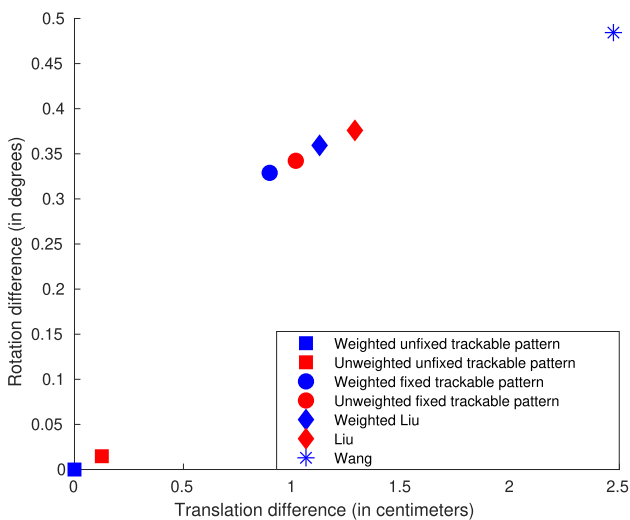


Figure 8: The calibration difference with regard to different setups and different methods.

- Intelligent Vehicles Symposium, 2005. Proceedings. IEEE.* IEEE, 2005, pp. 106–111.
7. B. Petit, J.-D. Lesage, C. Menier, J. Allard, J.-S. Franco, B. Raffin, E. Boyer, and F. Faure, “Multicamera real-time 3d modeling for telepresence and remote collaboration,” *International journal of digital multimedia broadcasting*, vol. 2010, 2010.
 8. S. Nair, G. Panin, M. Wojtczyk, C. Lenz, T. Friedelhuber, and A. Knoll, “A multi-camera person tracking system for robotic applications in virtual reality tv studio,” in *Proceedings of the 17th IEEE/RSJ International Conference on Intelligent Robots and Systems 2008*. 2008.
 9. T. Strauß, J. Ziegler, and J. Beck, “Calibrating multiple cameras with non-overlapping views using coded checkerboard targets,” in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE, 2014, pp. 2623–2628.
 10. R. Xia, M. Hu, J. Zhao, S. Chen, Y. Chen, and S. Fu, “Global calibration of non-overlapping cameras: state of the art,” *Optik-International Journal for Light and Electron Optics*, vol. 158, pp. 951–961, 2018.
 11. J. Wang, L. Wu, M. Q.-H. Meng, and H. Ren, “Towards simultaneous coordinate calibrations for cooperative multiple robots,” in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 410–415.
 12. P. Wunsch, S. Winkler, and G. Hirzinger, “Real-time pose estimation of 3d objects from camera images using neural networks,” in *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, vol. 4. IEEE, 1997, pp. 3232–3237.

Bionotes



Zaijuan Li

Technische Universität Darmstadt,
Darmstadt, Germany
zaijuan.li@mr.tu-darmstadt.de

Zaijuan Li received the M.Sc. degree in Electromechanical Engineering with major in Robotics from Harbin University of Science and Technology, Harbin, China, in 2012. She is currently working toward the Dr.-Ing. degree in the area of computer vision with the Control Methods and Robotics Laboratory, TU Darmstadt, Darmstadt, Germany. Her main research interests are in the field of multi-camera calibration and cooperative mobile vision systems as well as multi-robot localization.



Volker Willert

Technische Universität Darmstadt,
Darmstadt, Germany
vwillert@mr.tu-darmstadt.de

Volker Willert received the Dipl.-Ing. degree in electrical engineering and information technology and the Dr.-Ing. degree in control theory and robotics, with a focus on dynamical computer vision, from TU Darmstadt, Darmstadt, Germany, in 2002 and 2006, respectively. From 2005 to 2009, he was a Senior Scientist at Honda Research Institute Europe GmbH. Since July 2009, he has been with the Chair of the Control Methods and Robotics Laboratory, TU Darmstadt, and heads the research group Machine Vision and Autonomous Systems. His main research interests are in the fields of machine intelligence, computer vision, distributed controls, and machine learning for mobile robotics, multiagent systems, and driver assistance systems.