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

Zaijuan Li[1] and Volker Willert[1]

TU Darmstadt, Control Methods and Robotics Lab,
Landgraf-Georg Straße 4, 64283 Darmstadt

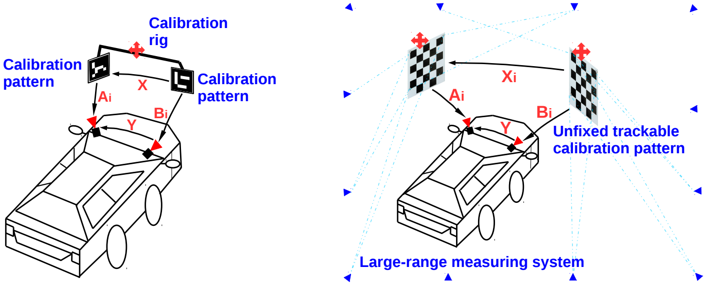
Abstract The calibration of the relative pose between rigidly connected cameras with non-overlapping fields of view (FOV) is a prerequisite for many applications. In this paper, we focus on the subtleties of experimental realization of such a calibration optimization method presented in [1]. We evaluate two strategies to adapt a given optimization process to find better local minima. The first strategy is the introduction of a quality measure for the image data used for calibration, which is based on the projection size of 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 second strategy to further improve accuracy is a careful data acquisition of pose pairs used for the calibration. We integrate the above strategies into different setups and demonstrate the improvement both in simulation and real-world experiment.

Keywords Quality measure, extrinsic calibration.

1 Introduction

Extrinsic camera calibration comprises the estimation of the relative pose between cameras with non-overlapping FOV. Especially, in the car industry, the calibration of cameras mounted on the car with vastly different FOV is ubiquitous [2], [3].

Different classifications of the existing calibration methods, together with detailed analyses and discussion of the methods could be found in [1], [4]. In principle, our proposed strategies are suitable for any setup that builds an objective function based on the reprojection error of 3D-2D point correspondences constrained by 3D-3D closed-loop pose transformation $\mathbf{AX} = \mathbf{YB}$. Here, we restrict ourselves to the following two setups given in Fig. 1(a) and Fig.1(b).



(a) Liu setup [5]

(b) Large-range measuring device based setup

Figure 1: Illustration of two setups that could apply our 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.

The first setup presented in [5] uses a movable calibration rig that rigidly links two planar calibration patterns whose relative pose is unknown (see Fig.1(a)). By changing the pose of the calibration rig relative to the camera, the initial estimation of $\mathbf{X}$ and $\mathbf{Y}$ can be recovered by solving $\mathbf{AX} = \mathbf{YB}$. The initial values are then applied to minimize the objective function based on the reprojection errors from all measurements. Similar to the first setup, the second setup in [1] (see Fig.1(b)) introduces high accuracy tracking system 'OptiTrack', so the two pattern boards used for recovering the relative pose to camera pairs could be accurately localized within the tracking system after aligning their coordinate frames with that of the tracking targets attached to them. In this case, the 3D-3D closed-loop pose transformation is formulated as $\mathbf{AX}_i$= $\mathbf{YB}$, where $\mathbf{X}_i$ is obtained from the tracking system. The extrinsic could thus be solved and 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 calibration, which is a big advantage, especially for mobile vehicles. However, the limited pose change of the calibration targets could result in instability [4]. Applying a large-range measuring system in the second setup is generally more accurate but the setup complexity and the costs are high.

2 Problem statement and optimization strategies

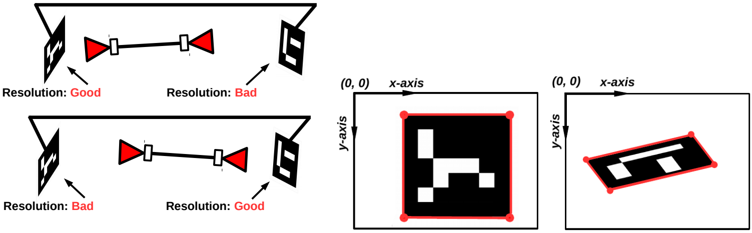
In [1], the optimization problem of Liu's setup is formulated as follows:

$$(\hat{\mathbf{R}}_X, \hat{\mathbf{t}}_X, \hat{\mathbf{R}}_Y, \hat{\mathbf{t}}_Y) = \underset{\mathbf{R}_X, \mathbf{t}_X, \mathbf{R}_Y, \mathbf{t}_Y}{\arg\min} \sum_{i=1}^{n} (\sum_{j=1}^{m} \|\epsilon_{ij}^{A}\|_2^2 + \sum_{l=1}^{o} \|\epsilon_{il}^{B}\|_2^2), \tag{1}$$

where $\mathbf{R}_X$, $\mathbf{t}_X$, $\mathbf{R}_Y$, $\mathbf{t}_Y$ are the unknown rotational and translational matrix to be optimized, i.e. the relative pose between two calibration patterns $\mathbf{X}$ and the camera pair $\mathbf{Y}$. m and o stand for the amount of the fiducial features from corresponding patterns, and n is the number of the collected pose pairs. ϵ_{ij}^{A} and ϵ_{il}^{B} are the reprojection errors from different camera frames. 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 paper, we use initial values for $\mathbf{X}$ and $\mathbf{Y}$ applying the method in [6] beforehand.

For each measurement pair $(\mathbf{A}_i, \mathbf{B}_i)$, both markers have to be in the FOV of the cameras such that all the coordinates of the fiducial feature projections can be extracted without outliers and with a certain accuracy. In practice, collecting a proper set of measurement pairs is challenging because of the rigid coupling of the patterns. With the assistance of the customized calibration device, a minor change in pose $\mathbf{A}_i$ or $\mathbf{B}_i$ would lead to an unpredictable change in the other, which indicates the hardness of capturing both calibration patterns with high resolution.

Fig.2(a) demonstrates the relationship between the calibration rig pose relative to the camera pair and the captured resolution quality. In Fig.2(b) an example is given to further explain this problem, which shows very *imbalanced* projection sizes of two different projec-



(a) Relationship between the pose and the resolution quality.

(b) An example of two fiducial patterns captured from different cameras.

Figure 2: 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. In this example, the left image gives a larger projection size of the pattern, hence higher resolution than the right image. When both images are corrupted by the same level noise, the left one is less sensitive to noise and produces a better pose estimation.

tions within one pose pair measurement. This discourages the objective function from including all the measurements and treating them equally.

Considering the unpleasant *imbalance* of the measurement quality which leads to the diversity of the projection size within one measurement pair, we introduce the additional weightings $\lambda_i^{A/B}$ to the objective (1) that are proportional to the projection size of the planar patterns, which leads to:

$$(\hat{\mathbf{R}}_X, \hat{\mathbf{t}}_X, \hat{\mathbf{R}}_Y, \hat{\mathbf{t}}_Y) = \operatorname*{arg\,min}_{\mathbf{R}_X, \mathbf{t}_X, \mathbf{R}_Y, \mathbf{t}_Y} \sum_{i=1}^{n} (\lambda_i^A \sum_{j=1}^{m} \|\epsilon_{ij}^A\|_2^2 + \lambda_i^B \sum_{l=1}^{o} \|\epsilon_{il}^B\|_2^2). \quad (2)$$

The weighting λ_i^A used for the reprojection error related with $\mathbf{A}_i$ is chosen to be the square root of the projection size $S(\mathbf{B}_i)$ related with reprojection $\mathbf{B}_i$ normalized by the full image size S_{max} and vice versa:

$$\lambda_i^A = \sqrt{S(\mathbf{B}_i)/S_{max}}, \quad (3)$$

$$\lambda_i^B = \sqrt{S(\mathbf{A}_i)/S_{max}}, \quad (4)$$

The reason for choosing such weighting lies in the replacement of $\mathbf{A}_i$ with $\mathbf{Y}\mathbf{B}_i\mathbf{X}^{-1}$. The reprojection error produced from $\mathbf{A}_i$ now depends on its replacement $\mathbf{Y}\mathbf{B}_i\mathbf{X}^{-1}$ which has the measurement $\mathbf{B}_i$ inside, so the quality of $\mathbf{B}_i$ influences the reprojection error of $\mathbf{A}_i$: $\mathbf{B}_i$ with good-quality should have more influence on the optimization and this leads to a higher weight λ_i^A. The same happens with the replacement of $\mathbf{B}_i$.

Except for the measurement quality, another major practical issue to reach accurate calibration results is a proper set of measurement pairs covering the six degrees of freedom of the poses $\mathbf{X}$ and $\mathbf{Y}$, which indicates the spatial distribution of pose pairs also influences the calibration results. This requires that images of the calibration objects should be taken from as many different poses as possible. However, generating a pose pair set with good-quality and comparatively scattered spatial distribution which are essential for accurate calibration is subtly tricky. Therefore, instead of including all collected pose pairs, our second strategy picks out a subset with comparatively scattered pose pairs and good measurements. In this case, compromises have to be made between pose change variety and measurement quality.

3 validation on simulated dataset

3.1 Synthetic Dataset

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

To generate the synthetic dataset, an exhaustive searching program is first run based on known ground truth and camera intrinsic parameters to produce a 'pose pair bank' which consists of over 12,000 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 span the whole possible measurement space; The projection size of the calibration object generated by the corresponding pose pair must exceed a certain threshold, which guarantees the minimum quality of the measurements. In this experiment, the threshold is set to 0.14 of the full image plane. The synthetic measurements are then generated based on 'pose pair bank': First, the true pose pairs are randomly extracted from the bank; The noise-free 2D coordinates obtained through the projection process are then corrupted with Gaussian noise; In the end, the noise-corrupted 2D coordinates are used to recover the noisy pose pairs which will be taken as the measurements of $\mathbf{A}_i$ and $\mathbf{B}_i$.

3.2 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 the $\mathbf{X}$ is taken as the example.

We apply the method of Wunsch et al. in [7] 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 $\mathbf{e}_X^R$ is then defined as:

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

in which $' \cdot '$ denotes the inner product of two quaternion vectors. Here the rotation error is represented by the angles returned by $arccos$ and mapped to $[0, 90°]$.

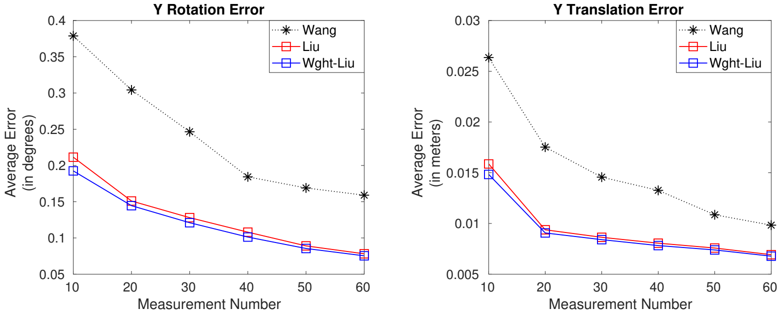
The estimated translation vector is described as $\hat{\mathbf{t}}_X$, and the ground truth is $\mathbf{t}_X$. The translation error is computed as follows,

$$\mathbf{e}_X^t = \|\mathbf{t}_X - \hat{\mathbf{t}}_X\|. \tag{6}$$

3.3 Simulation results

The results from the method proposed in [5] as well as after applying our strategies will be compared. In addition, the calibration results of the method in [6] are also presented since the above methods take its estimation of $\mathbf{X}$ and $\mathbf{Y}$ as initial values. Considering the improvement brought by the second strategy is not noticeable and would cover the other results, in what follows we demonstrate the benefits of strategies individually. For the validation of the proposed quality measure factor, those methods are labeled as Liu, 'Weighted Liu's method' (Wght-Liu), and Wang.

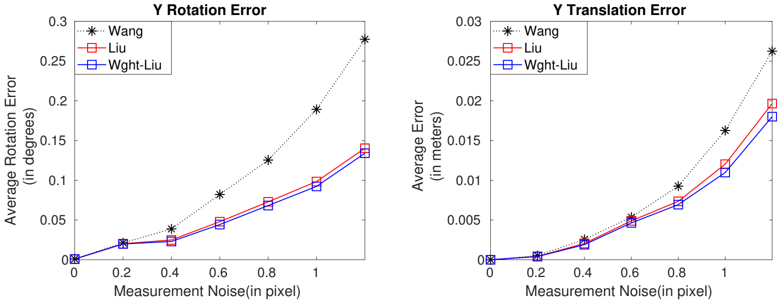
In the first setting, the measurement number changes from 10 to 60 with the fixed Gaussian noise of 1.0 pixel; In the second one, the added Gaussian noise varies from 0.2 to 1.2 pixels with a fixed number of 40 measurement pairs. The results shown below are taken the average of 100 iteration runs. For each iteration, the measurements are extracted from the 'pose pair bank' and processed applying each method. The calibration results are the average of overall calibration errors. The code for the calibration model as well as the optimization method is available online[1].



(a) Translation error of $\mathbf{Y}$.

(b) Rotation error of $\mathbf{Y}$.

Figure 3: Estimation error with regard to increased number of measurements and different methods.



(a) Translation error of $\mathbf{Y}$.

(b) Rotation error of $\mathbf{Y}$.

Figure 4: Estimation error with regard to increased image Gaussian noise and different methods.

Fig.3 and Fig.4 demonstrate the estimation error of $\mathbf{Y}$ under different

[1] https://github.com/zaijuan/eye-to-eye-calibration.git

settings. Since the comparison 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 our weighting factor strategy: Wght-Liu after applying the weighting factor gives the best results regardless of different settings; With the increase of noise level, the benefits from applying the weighting factor become more noticeable.

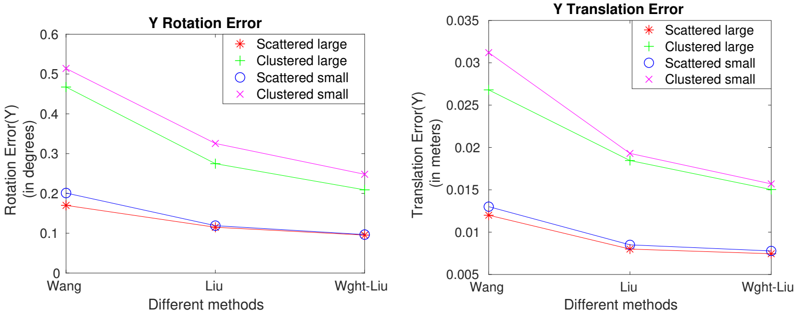
To demonstrate the importance of the spatial distribution of pose pairs to the calibration results, measurement sets with the following characteristics could be generated from 'pose pair bank': a). Spatially scattered pose pair set with larger projection size. b). Spatially clustered pose pair set with larger projection size. c). Scattered distributed pose pair set with smaller projection size. d). Clustered distributed pose pair set with smaller projection size. The 'scattered' and 'clustered' spatial distributions are comparative concepts. 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. Since each calibration object has to be placed in the FOV of its corresponding camera, the measurement space is reduced, so the pose pairs in the scattered set have comparatively larger while still limited differences in translation and rotation. Same for the measurement quality. The bad measurements are somewhat bad only when compared to the good ones, they are still guaranteed the minimum required quality.

We use those four extreme types of measurement sets to emphasize the improvement after choosing comparatively scattered pose pair distribution. For each configuration, the measurement number is set to be 40, and the noise level is 1.0. Fig.5 demonstrates the calibration results of different methods with different types of measurement sets.

Same as before, a final bundle adjustment with weighting factors is applied to refine calibration results. Since the relative pose $\mathbf{X}_i$ between the calibration patterns for each measurement is known, the replacement of $\mathbf{A}_i$ and $\mathbf{B}_i$ now becomes:

$$\mathbf{A}_i = \mathbf{Y}\mathbf{B}_i\mathbf{X}_i^{-1}, \tag{7}$$

$$\mathbf{B}_i = \mathbf{Y}^{-1}\mathbf{A}_i\mathbf{X}_i. \tag{8}$$



(a) Rotation error of $\mathbf{Y}$.

(b) Translation error of $\mathbf{Y}$.

Figure 5: Estimation error with regard to different pose pair configurations and different methods.

For all methods, spatially scattered pose pairs with better measurement quality (larger projection size) generate the best accurate estimation, while the clustered spatially pose pairs with smaller projection size result in the worst estimation results. When the measurements have the same quality level, scattered pose pairs always produce better results than clustered ones, which implies the calibration methods are more demanding on the distribution of pose pairs than their measurement quality.

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

4 Real experimental results

4.1 Experiment setup

In the real experiment, both Liu's setup and the setup in [1] are implemented. Those experiments are carried out as follows:

In the first experiment, the calibration rig with two rigidly linked calibration patterns is placed to a variety of poses relative to the camera pairs and the corresponding pictures containing the calibration pattern are taken to recover the relative pose between them. In the end, a measurement set containing different pose pairs $\mathbf{A}_i$ and $\mathbf{B}_i$ is generated and used for different calibration methods.

With the assistance of the tracking system, the unknown relative pose $\mathbf{X}$ between the calibration objects could be accurately recovered. We name this configuration 'fixed trackable pattern'. In this case, the collected $\mathbf{A}_i$, $\mathbf{B}_i$, and the recovered $\mathbf{X}$ are used to run a final bundle adjustment (9) including weighting factors similar to (2) to refine calibration results:

$$(\hat{\mathbf{R}}_Y, \hat{\mathbf{t}}_Y) = \underset{\mathbf{R}_Y, \mathbf{t}_Y}{\arg\min} \sum_{i=1}^{n} (\lambda_i^A \sum_{j=1}^{m} \|\epsilon_{ij}^A\|_2^2 + \lambda_i^B \sum_{l=1}^{o} \|\epsilon_{il}^B\|_2^2). \tag{9}$$

The further improvement brought by the tracking system is that the two calibration patterns do not have to be rigidly linked. This extra flexibility facilitates the improvement of measurement quality since each calibration pattern could be placed into positions relative to corresponding cameras which generate the best possible estimates. So in the second experiment (setup), two pattern boards are independently placed to different poses relative to cameras. We describe this configuration as 'unfixed trackable pattern'. A set of $\mathbf{A}_i$ and $\mathbf{B}_i$ with better measurement quality together with the corresponding ground truth of $\mathbf{X}_i$ obtained from the tracking system is collected.

Same as before, a final bundle adjustment with weighting factors is applied to refine calibration results. Since the relative pose $\mathbf{X}_i$ between the calibration patterns for each measurement is known, the replacement of $\mathbf{A}_i$ and $\mathbf{B}_i$ now becomes:

$$\mathbf{A}_i = \mathbf{Y}\mathbf{B}_i\mathbf{X}_i^{-1}, \tag{10}$$

$$\mathbf{B}_i = \mathbf{Y}^{-1}\mathbf{A}_i\mathbf{X}_i. \tag{11}$$

4.2 Experimental results

We use the same error criteria to evaluate the calibration difference of different methods. The benchmark is set as the weighted estimation of configuration ‘unfixed trackable pattern’. We use the term ‘difference’ 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.

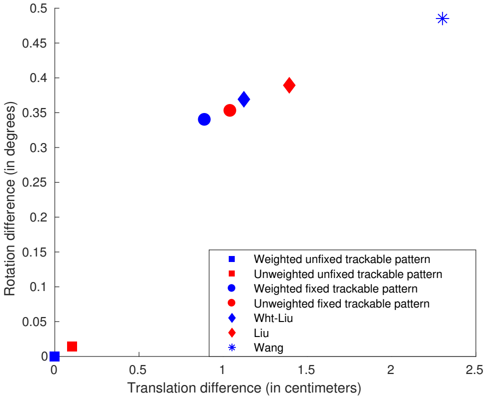


Figure 6: Calibration difference of $\mathbf{Y}$ with regard to different configurations and different methods.

Fig. 6 shows the calibration differences of different setups and different methods. Although in ‘unfixed trackable pattern’ configuration, the difference between weighted and unweighted estimation is minor since all patterns are captured with relatively high resolution, we could still verify that applying weighting factor generates less deviation compared to benchmark regardless of the configuration.

5 Conclusion and future work

In this paper, we discussed the subtleties of certain calibration methods and proposed two optimization strategies applicable to calibration setups that could minimize the reprojection error of 3D-2D point correspondences constrained by rigid 3D-3D closed-loop pose transformations $\mathbf{AX} = \mathbf{YB}$. First, we introduce an additional quality measure factor to the objective function, which helps enlarge the measurement space and improves the calibration accuracy. Hence, instability could be alleviated and the robustness could be safely guaranteed. Besides, by carefully choosing a measurement subset the possibility of getting trapped in a worse local minimum is reduced. In future work, we will focus on refining the weighting factor, which is now simply based on the projection area of the calibration pattern.

References

1. 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.
2. S. Dabral, S. Kamath, V. Appia, M. Mody, B. Zhang, and U. Batur, "Trends in camera based automotive driver assistance systems (adas)," in *Circuits and Systems (MWSCAS), 2014 IEEE 57th International Midwest Symposium on*. IEEE, 2014, pp. 1110–1115.
3. D. Cheda, D. Ponsa, and A. Lóz, "Camera egomotion estimation in the adas context," in *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, 2010, pp. 1415–1420.
4. 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.
5. 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.
6. 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.
7. 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.