Global Uncertainty-based Selection of Relative Poses for Multi Camera Calibration

Ferid Bajramovic, Joachim Denzler
Chair for Computer Vision, Friedrich-Schiller-University Jena
Ernst-Abbe-Platz 2, 07743 Jena, Germany
{bajramov,denzler}@informatik.uni-jena.de
http://www.inf-cv.uni-jena.de

Abstract

Extrinsically calibrating a multi camera system from scene images is in general a very difficult problem. One promising approach uses pairwise relative poses as input. As a limited number of relative poses suffices, we propose automatically selecting only the most reliable ones. We present theoretically sound local and global uncertainty measures on relative poses and a selection criterion based on these measures. We show that our criterion is equivalent to computing a shortest subgraph consisting of shortest triangle connected paths and provide an efficient algorithm. In experiments on synthetic and real data, we show that our selection algorithm produces greatly improved calibration results, both, in case of varying portions of outliers as well as varying noise.

1 Introduction

Multi-camera systems become increasingly important in computer vision. For many applications, however, the system has to be calibrated, i.e. the intrinsic and relative extrinsic parameters of the cameras have to be determined. Existing methods can be very roughly classified by the type of input or scene knowledge they require: 1. Pattern based: a classical (planar) calibration pattern either has to be visible in all images [16] or the poses of multiple calibration objects have to be known [7]. 2. LED based: some easily detectable, moving single feature, like an LED in a dark room, is recorded over time [1, 14, 4, 2]. 3. Self-calibration (in the broader sense): images are taken from an unknown scene, typically with some (unknown) 3D structure and texture [12, 15]. Clearly, the third class is most appealing from a practical point of view, as it is the most flexible one.

From a pure multiple view geometry point of view, multiple camera calibration can be interpreted as a structure from motion (sub)problem [6]. In case of self-calibration of multiple physical cameras, however, point correspondences have to be extracted from images in a medium or wide baseline situation, which typically leads to a very high portion of outliers. In this paper, we focus on such a situation. We want to stress that we make no assumptions about the scene except that it has some arbitrary 3D structure and texture. As the relative pose of two cameras can be estimated even in presence of very many outliers [13, 3], we will use known relative poses between some camera pairs as input. The quality of the results, of course, highly depends on the relative pose estimates. This problem has been briefly mentioned by Chen, David and Slusallek [4]. It has been

BMVC 2008 doi:10.5244/C.22.74

more recently addressed by Martinec and Pajdla [12, 11]. They weight relative poses by a measure based on the number of inliers found by RANSAC and also on the "importance" of a relative pose. While their measure is plausible, a theoretical justification is missing.

A further method has been proposed by Vergés-Llahí, Moldovan and Wada [15]. They use an uncertainty measure for relative poses, which consists of a residual and a constraint violation term. Note, however, that constraint violations can be avoided by using the five point algorithm [13]. Selection of relative poses is performed by finding triangle connected shortest paths in a graph connecting cameras according to known relative poses with edge weights set to these uncertainty values. We adopt their graph-based approach.

In this paper, we make the following contributions: We propose three theoretically sound local and global uncertainty measures on relative poses. As a basis, we use the work of Engels and Nistér [5], but suggest using the Blake-Zisserman distribution instead of the Cauchy distribution. Based on these measures, we formulate a sound selection criterion for relative poses. We provide a theoretical justification for the triangle connected shortest paths approach [15] and present an efficient algorithm. We also formulate the choice of the reference camera (pair) as an optimization problem, which completes our selection criterion, and also provide an efficient algorithm for this problem. Experiments on synthetic and real data show a great improvement of calibration results by our selection criterion compared to a naïve approach in presence of varying noise as well as varying outlier portions.

The paper is structured as follows: Section 2 presents the specific problem treated in this paper and introduces important notation. Section 3 continues with the theoretical and algorithmic contributions of the paper. In section 4, we present systematic experiments on simulated and real data. Conclusions are given in section 5.

2 Basics

2.1 Relative and Absolute Poses

We assign a number $i \in \{1,\dots,n\}$ to each camera and denote the set of all cameras by $\mathcal{V} = \{1,\dots,n\}$. The relative pose between two cameras i and j is denoted by $(\boldsymbol{R}_{ij}, \boldsymbol{t}_{ij})$. It is defined such that it maps a 3D point $\boldsymbol{p}_i$ from the ith camera coordinate system to the jth one as follows: $\boldsymbol{p}_j = \boldsymbol{R}_{ij}\boldsymbol{p}_i + \boldsymbol{t}_{ij}$. The absolute poses (transforming points from world coordinates to kth camera coordinates) are denoted by $(\boldsymbol{R}_k, \boldsymbol{t}_k)$. Note that relative poses have a double subscript, while absolute poses have a single one. When estimating relative poses from images only, the scale of $\boldsymbol{t}_{ij}$ cannot be determined. We emphasize unknown scale by a star: $\boldsymbol{t}^*_{ij}$. Relative pose estimation will be treated in section 3.1. A multi camera system together with the set of all known relative poses can be represented by the *camera dependency graph* [15] $\mathcal{G}_\mathrm{R} = (\mathcal{V}, \mathcal{E})$: each camera is a vertex and camera i is connected to camera j iff the relative pose $(\boldsymbol{R}_{ij}, \boldsymbol{t}^*_{ij})$ is known. An example is given in Figure 1 (left). Note that edges are directed to distinguish $(\boldsymbol{R}_{ij}, \boldsymbol{t}^*_{ij})$ from $(\boldsymbol{R}_{ji}, \boldsymbol{t}^*_{ji})$. As relative poses can be easily inverted, however, all edges will be treated as bidirectional.

2.2 Calibration Task from Relative Poses

We assume known relative poses up to scale $(\boldsymbol{R}_{ij}, \boldsymbol{t}^*_{ij})$ and seek to compute absolute poses $(\boldsymbol{R}_k, \boldsymbol{t}_k)$ expressed in an arbitrary (but common) world coordinate system up to only one

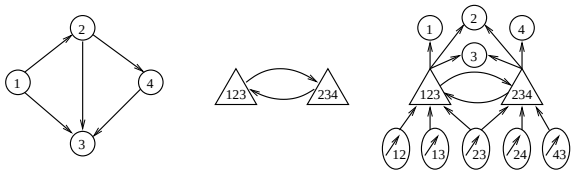

Figure 1: From left to right: a camera dependency graph with four cameras and five known relative poses, the associated triangle graph, and the extended triangle graph.

common, unknown scale factor—in other words, we want to calibrate the multi camera system up to a 3D similarity transformation. More precisely, calibration is possible up to one 3D similarity transformation per triangle connected component in the camera dependency graph [10]. We will assume a triangle connected camera dependency graph. Otherwise, the algorithms can be applied to each component separately.

2.3 Calibration via Triangulation

A *triangle* in $\mathcal{G}_\mathrm{R}$ consists of three cameras and three known relative poses. If one of the scales $\|\boldsymbol{t}_{ij}\|$ is known (or fixed), the remaining two scales can be computed by triangulation [6]. By treating one triangle after the other such that all consecutive triangles have a common edge, all scales can be computed up to one common unknown scale parameter. The calibration algorithm can be formulated nicely by using the *triangle graph* $\mathcal{G}_\mathrm{T} = (\mathcal{T}, \mathcal{E}_\mathrm{T})$, which is defined as follows (Figure 1, center): each triangle in $\mathcal{G}_\mathrm{R}$ becomes a vertex in $\mathcal{G}_\mathrm{T}$, two such vertices are connected iff the two triangles have a common edge in $\mathcal{G}_\mathrm{R}$. Once the scales have been estimated, the absolute poses can be extracted from the relative poses based on the following two equations: $\boldsymbol{R}_j = \boldsymbol{R}_{ij}\boldsymbol{R}_i$ and $\boldsymbol{t}_j = \boldsymbol{R}_{ij}\boldsymbol{t}_i + \boldsymbol{t}_{ij}$. These steps can be integrated into the following *relative pose calibration algorithm*: First choose a starting triangle $(i, j, k) \in \mathcal{T}$ and set $\|\boldsymbol{t}_{ij}\| = 1$, $\boldsymbol{R}_i = \boldsymbol{I}$, and $\boldsymbol{t}_i = 0$. Then traverse $\mathcal{G}_\mathrm{T}$, e.g. via breadth first search (BFS). When visiting a triangle, estimate the missing scale factors via triangulation and extract the missing absolute poses.

3 Selection of Relative Poses

In the relative pose calibration algorithm, the traversal order is unspecified. Furthermore, in case of a dense camera dependency graph, only a rather small subset of triangles has to be visited in order to calibrate all cameras. Thus, the traversal order corresponds to a selection of triangles and thus relative poses. Vergés-Llahí, Moldovan and Wada [15] calibrate along shortest triangle connected paths in the camera dependency graph. They define the weight of a relative pose edge by an uncertainty measure. We adopt that idea.

An uncertainty measure should not only capture the local precision of a relative pose estimate, but also the global uncertainty caused by ambiguities. Furthermore, it should be well suited for a theoretically sound formulation of relative pose selection as an optimization problem. The uncertainty measure of Vergés-Llahí, Moldovan and Wada [15] does not seem to satisfy these criteria. Engels and Nistér [5] proposed a sampling based approach to estimate the global uncertainty of a relative pose estimate. Base on that, we will define three uncertainty measures.

3.1 Uncertainty Measures

First, we will briefly summarize their approach in a slightly different formulation (focussing on the five point variant). Let $\mathcal{D}$ denote the set of all point correspondences. The posterior probability density function of a relative pose $(\boldsymbol{R}, \boldsymbol{t}^*)$ is denoted by $p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D}) \propto p(\mathcal{D} \mid \boldsymbol{R}, \boldsymbol{t}^*) p(\boldsymbol{R}, \boldsymbol{t}^*)$. In the experiments, we assume a flat prior $p(\boldsymbol{R}, \boldsymbol{t}^*)$. Very similar to Engels and Nistér [5], we define the data likelihood based on the Cauchy distribution as follows:

$$p(\mathcal{D} \mid \boldsymbol{R}, \boldsymbol{t}^*) \propto \left(\prod_{\boldsymbol{d} \in \mathcal{D}} \frac{\alpha}{\alpha^2 + s(\boldsymbol{R}, \boldsymbol{t}^*, \boldsymbol{d})} \right)^{|\mathcal{D}|^{-\phi}} , \tag{1}$$

where α is a scale parameter, which is typically set to 1, ϕ widens peaks of the distribution ($\phi = 0.5$ is recommended, $\phi = 0$ corresponds to assuming independent point correspondences), and $s(\boldsymbol{R}, \boldsymbol{t}^*, \boldsymbol{d})$ denotes the Sampson epipolar error [5, 6], which approximates the squared reprojection error of the point correspondence $\boldsymbol{d}$ and the relative pose $(\boldsymbol{R}, \boldsymbol{t}^*)$ together with the known intrinsic calibration.

The density over translation directions is given by $p(\boldsymbol{t}^* \mid \mathcal{D}) = \int p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D}) d\boldsymbol{R}$. Engels and Nistér [5] approximate the integral using Laplace's method. Instead, we use the even simpler approximation $p(\boldsymbol{t}^* \mid \mathcal{D}) \propto \max_{\boldsymbol{R}} p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D})$, which equals the Laplace approximation if $\det(\nabla^2_{\boldsymbol{R}} \log p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D}))$ is independent of $\boldsymbol{t}^*$ [9].

The density $p(\boldsymbol{t}^* \mid \mathcal{D})$ is represented discretely over the range of $\boldsymbol{t}^*$ by a $c \times c$ matrix $\boldsymbol{A}$. The idea consists of generating relative pose hypotheses by sampling, computing the according indices (a, b) into $\boldsymbol{A}$ and keeping the maximum for each entry of $\boldsymbol{A}$. As a byproduct, an estimate for $\boldsymbol{t}^*$ can be computed, which we denote by $\hat{\boldsymbol{t}}^*$ in this section: $\hat{\boldsymbol{t}}^* \approx \operatorname{argmax}_{\boldsymbol{t}^*} p(\boldsymbol{t}^* \mid \mathcal{D})$. The following algorithm contains further details:

1. Initialize $\boldsymbol{A} = 0$ and $\gamma = 0$ (storing the maximum occuring value of $p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D})$).
2. Repeat m times:
   (a) Draw a sample from $\mathcal{D}$ and apply the five point algorithm [13, 3] to estimate the relative pose $(\boldsymbol{R}, \boldsymbol{t}^*)$.
   (b) Normalize $\boldsymbol{t}^*$ to length 1, choose the sign such that $\boldsymbol{t}^*$ is in the upper unit-hemisphere, and map the first two coordinates to discrete matrix indices (a, b).
   (c) Set $\boldsymbol{A}(a, b) = \max(\boldsymbol{A}(a, b), p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D}))$.
   (d) If $\gamma > p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D})$, set $\gamma = p(\boldsymbol{R}, \boldsymbol{t}^* \mid \mathcal{D})$ and $\hat{\boldsymbol{t}}^* = \boldsymbol{t}^*$.
3. Normalize $\boldsymbol{A}$ such that $\sum_{a,b} \boldsymbol{A}(a, b) = 1$.

We choose a rather low resolution of $c = 100$ for $\boldsymbol{A}$ and only perform $m = 10000$ sampling iterations. Nevertheless, this still gives a reasonable approximation to $p(\boldsymbol{t}^* \mid \mathcal{D})$.

As the experiments will show, equation (1) does not work well in case of a large amount of outliers. As an alternative, we propose the following data likelihood based on the Blake-Zisserman "distribution"[1] [6]:

$$p(\mathcal{D} \mid \boldsymbol{R}, \boldsymbol{t}^*) \propto \left(\prod_{\boldsymbol{d} \in \mathcal{D}} \left(\exp\left(-\frac{s(\boldsymbol{R}, \boldsymbol{t}^*, \boldsymbol{d})}{\sigma^2} \right) + \epsilon \right) \right)^{|\mathcal{D}|^{-\phi}} , \tag{2}$$

[1] If we make the reasonable assumption that outlier points only occur within the image area, the uniform part of the distribution can be limited to finite support and we actually get a valid probability density function.

where σ^2 is the variance of the Gaussian component (we set $\sigma = 1$) and ϵ defines the relative weight of the uniform component (we set $\epsilon = 0.0002$).

We investigate the following uncertainty measures $\omega(\hat{\boldsymbol{t}}^*)$ based on $p(\boldsymbol{t}^* \mid \mathcal{D})$, which is approximated by $\boldsymbol{A}$:

1. The local *information* measure $\omega_{\mathrm{I}}(\hat{\boldsymbol{t}}^*) = -\log p(\hat{\boldsymbol{t}}^* \mid \mathcal{D})$.
2. The global *entropy* measure $\omega_{\mathrm{E}}(\hat{\boldsymbol{t}}^*) = -\int p(\boldsymbol{t}^* \mid \mathcal{D}) \log p(\boldsymbol{t}^* \mid \mathcal{D}) d\boldsymbol{t}^*$.
3. The *smoothed information* measure, defined as the information of the smoothed density: $\omega_{\mathrm{S}}(\hat{\boldsymbol{t}}^*) = -\log \int \mathcal{N}(\boldsymbol{t}^*; \hat{\boldsymbol{t}}^*, \boldsymbol{\Sigma}) p(\boldsymbol{t}^* \mid \mathcal{D}) d\boldsymbol{t}^*$, where $\mathcal{N}(\boldsymbol{t}^*; \hat{\boldsymbol{t}}^*, \boldsymbol{\Sigma})$ denotes the normal distribution with mean $\hat{\boldsymbol{t}}^*$ and covariance matrix $\boldsymbol{\Sigma}$ (we set $\boldsymbol{\Sigma} = \sqrt{5}\boldsymbol{I}$). Note that the integral can be interpreted as a variant of the confidence measure used by Engels and Nistér [5] (using a Gaussian kernel instead of a confidence area).

3.2 Criterion for the Selection of Relative Poses

Calibrating the pose of a camera j relative to a reference camera i involves a set $\mathcal{P} = \{(i_1, j_1), \ldots, (i_l, j_l)\}$ of relative poses such that $\mathcal{P}$ is a triangle connected path from i to j. As pointed out before, there are often several alternative sets $\mathcal{P}$. Based on the uncertainty measures defined above, we assign an uncertainty to each set $\mathcal{P}$ and choose the set with minimum uncertainty. Assuming independence, we have $p(\mathcal{P} \mid \mathcal{D}) = \prod_k p(\boldsymbol{t}^*_{i_k,j_k} \mid \mathcal{D})$. This leads to the following expression for all three uncertainty measures: $\omega(\mathcal{P}) = \sum_k \omega(\boldsymbol{t}^*_{i_k,j_k})$. As an example, we present the derivation for the information measure: $\omega_{\mathrm{I}}(\mathcal{P}) = -\log p(\mathcal{P} \mid \mathcal{D}) = -\log \prod_k p(\boldsymbol{t}^*_{i_k,j_k} \mid \mathcal{D}) = \sum_k -\log p(\boldsymbol{t}^*_{i_k,j_k} \mid \mathcal{D}) = \sum_k \omega_{\mathrm{I}}(\boldsymbol{t}^*_{i_k,j_k})$.

Choosing a suitable set $\mathcal{P}$ with minimum uncertainty is equivalent to finding a shortest triangle connected path in the camera dependency graph, where the weight of an edge (k, l) is defined as the uncertainty $\omega(\boldsymbol{t}^*_{k,l})$. This provides a theoretical justification for the shortest triangle connected paths approach [15].

Using this approach to calibrate a whole multi camera system involves shortest triangle connected paths from i to all other cameras. Note, however, that the union $\mathcal{U}$ of these paths also has to be triangle connected. While the common reference camera only guarantees that $\mathcal{U}$ is connected, using a common reference edge instead actually guarantees that $\mathcal{U}$ is *triangle* connected. Hence, we use shortest paths from a common relative pose to all cameras. We choose the reference edge $e \in \mathcal{E}$ with mimimum total uncertainty $\omega(\mathcal{U})$. In other words, we use the *shortest* triangle connected *shortest* paths subgraph. This extends the triangle connected shortest paths approach of Vergés-Llahí, Moldovan and Wada [15].

3.3 Computing Shortest Triangle Connected Paths

In this section, we present an efficient algorithm for our selection criterion. First, we construct the directed *extended triangle graph* $\mathcal{G}_{\mathrm{E}}$ as follows (Figure 1, right): Each triangle becomes a vertex. Additionally, we add an *entry vertex* for each relative pose and an *exit vertex* for each camera. Each entry vertex is connected to each triangle containing the respective relative pose. The weight of such an edge is the summed uncertainty of all relative poses in the triangle. Each triangle is also connected to the exit vertex of each camera it contains. These edges have weight zero. Each pair of triangles, which is connected in $\mathcal{G}_{\mathrm{T}}$, is connected in both directions. The weight of each edge is the summed

uncertainty of the *two* "new" relative poses in the target triangle (i.e. the common relative pose is ignored).

The graph is defined such that a shortest path from an entry vertex to an exit vertex corresponds to a shortest triangle connected path in $\mathcal{G}_\mathrm{R}$ with fixed first edge. We can thus use a standard shortest paths algorithm (e.g. Dijkstra) to compute the shortest paths from an entry vertex e to all exit vertices (*not* to *all* vertices in the graph!), and also the according shortest paths tree. This tree corresponds to a triangle connected subgraph $\mathcal{U}_e$ of the camera dependency graph, which contains triangle connected shortest paths from e to all cameras. We apply Dijkstra to all entry vertices e and choose the resulting subgraph $\mathcal{U}_e$ with minimum total uncertainty $\omega(\mathcal{U}_e)$.

Note that Vergés-Llahí, Moldovan and Wada [15] mention the need for an efficient shortest triangle connected paths algorithm, but do not actually present one. This gap is filled by our algorithm. Note, however, that we do not actually compute shortest triangle connected paths starting at a reference camera, but at a reference relative pose. In order to do the former, simply modify the extended triangle graph such that it has an entry vertex for each camera instead of each relative pose. Note also that the shortest (ordinary) paths algorithm in the related paper by Martinec and Pajdla [11] solves a *different* problem.

4 Experiments

As the multi camera system can only be calibrated up to a 3D similarity transformation, we compute and apply an optimal transformation between the calibration result and the ground truth using a linear algorithm followed by nonlinear refinement. As an error measure for the calibration of a multi camera system in comparison to the ground truth, we use the mean position error $e = \frac{1}{n}\sum_i \|\boldsymbol{R}_i\boldsymbol{t}_i - \boldsymbol{R}_{\mathrm{G},i}\boldsymbol{t}_{\mathrm{G},i}\|_2$, where the subscript G indicates ground truth. The ground truth calibration is normalized such that the distance of the first two cameras is one. This defines the scale of the error measure e.

4.1 Simulation

The simulation consists of six virtual pinhole cameras $(\boldsymbol{K}, \boldsymbol{R}_{\mathrm{G},i}, \boldsymbol{t}_{\mathrm{G},i})$ with image size 640×480 and calibration matrix $\boldsymbol{K} = ((1500, 0, 320), (0, 1500, 240), (0, 0, 1))^\mathrm{T}$. The scene consists of 100 random 3D points uniformly distributed in a cuboid. The cameras are placed in a circle above the points looking roughly towards the center of the cuboid such that all cameras can see all points. The height above the scene of every second camera is slightly increased to avoid planarity. The 3D points are projected into the cameras with subpixel precision. Noise is simulated by adding random values uniformly distributed in $[-\phi/2, \phi/2]$ to all coordinates. We choose $\phi = 1$ for all experiments. Outliers are simulated by replacing a certain fraction of the point correspondences of each camera pair by points uniformly distributed within the image areas. We perform experiments with the following amounts of outliers: 0%, 30%, 70%, and 80%. All experiments are repeated 50 times and the resulting mean position errors e are plotted in sorted order. This corresponds to computing all error quantiles. A good algorithm should, of course, have a low error curve. The curve should also be flat, which is a sign for highly repeatable results.

We perform the following two experiments. In the first one, we increase the portion of outliers of the camera pairs (1,2), (2,3), (3,4) and (4,5) such that only half of the inliers remain (e.g. instead of 30%, these camera pairs have 65% outliers). In the second

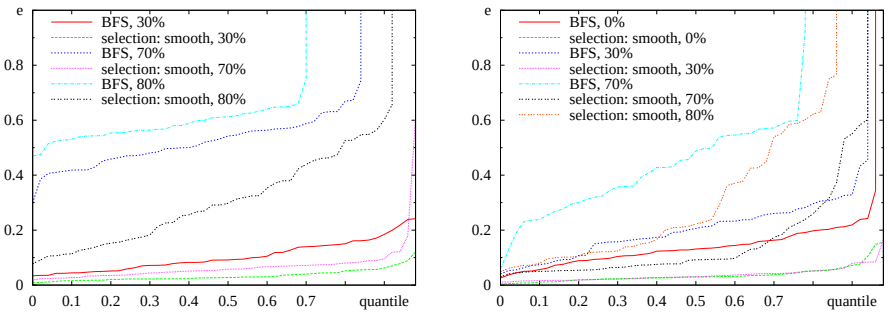

Figure 2: Comparison of the BFS traversal with our uncertainty based selection of relative poses using the smoothed information measure and the Blake-Zisserman distribution. The percent values denote the portion of outliers. The plot shows sorted mean position error values e (quantiles). Left: first experiment, right: second experiment.

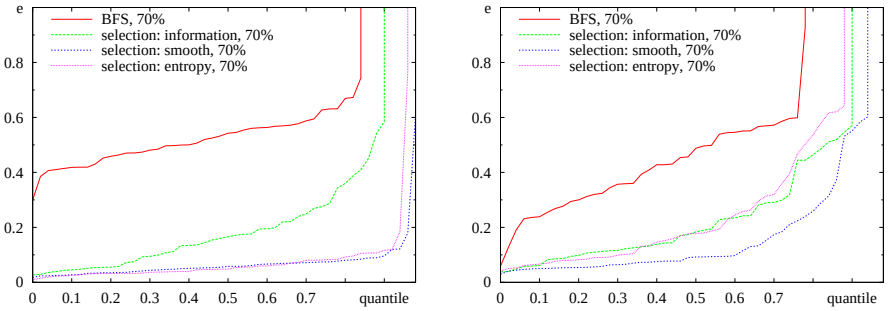

Figure 3: Comparison of the three uncertainty measures using the Blake-Zisserman distribution; including the BFS results. The percent values denote the portion of outliers. The plot shows sorted mean position error values e (quantiles). Left: first experiment, right: second experiment.

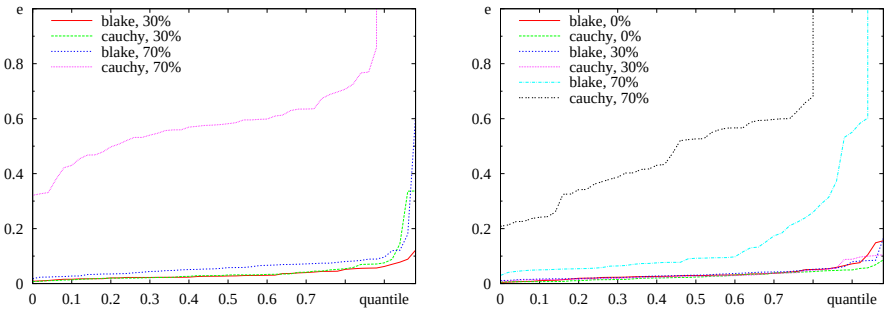

Figure 4: Comparison of the Cauchy and Blake-Zisserman distribution for the selection of relative poses using the smoothed information measure. The percent values denote the portion of outliers. The plot shows sorted mean position error values e (quantiles). Left: first experiment, right: second experiment.

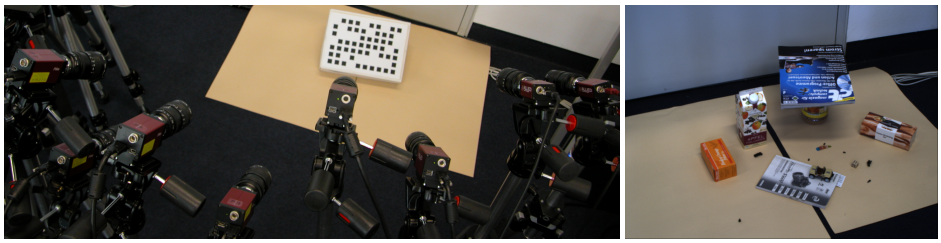

Figure 5: Setup of the real experiment. Left: the multi camera system observing the pattern for the Zhang calibration, right: the scene (image from the sixth camera).

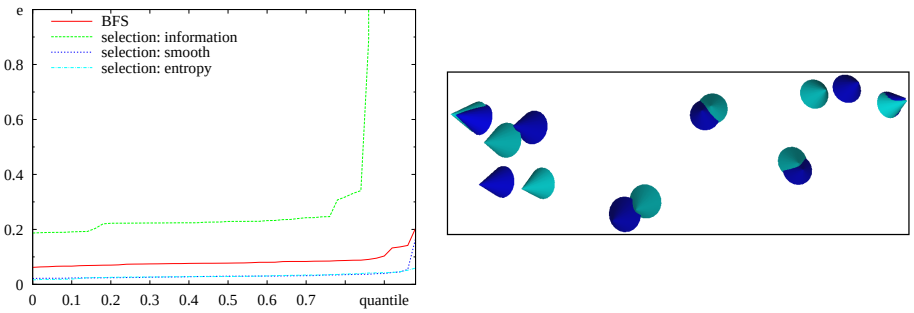

Figure 6: Experiment on real data. Left: mean position error quantile results for BFS as well as our selection criterion. Relative pose and uncertainty estimation used the Blake-Zisserman distribution. Right: a visualization of a calibration result (cyan / light gray) with a mean position error $e \approx 0.03$ compared to a Zhang calibration (blue / dark gray).

experiment, the noise level of these camera pairs is increased from 1 to 5 without changing the portion of outliers. We will compare our selection criterion to the BFS traversal and also compare the three uncertainty measures and the two probability density functions with each other.

First, we compare the BFS traversal to our selection criterion using the Blake-Zisserman distribution and the smoothed information measure. Figure 2 clearly shows that our selection criterion greatly improves the calibration results in both situations. This shows that in both cases the camera pairs with contaminated data can be identified by the smoothed information measure and that the selection mechanism is able to avoid them.

Figure 3 contains the comparison of the uncertainty measures in case of 70% outliers using the Blake-Zisserman distribution. Independent of the uncertainty measure, our selection criterion gives better results than BFS traversal. In the first experiment, the smoothed information measure and the entropy show a very similar performance. The purely local information measure is inferior. In the second experiment, the smoothed information measure is clearly better than the other two. The global entropy measure is obviously not very good at distinguishing situations with different noise levels. The smoothed information measure gives the best overall results by taking local *and* global uncertainty into account.

Finally, we present a comparsion of the two probability density functions in Figure 4.

Up to an outlier level of 30%, the Cauchy distribution works well and gives very similar results to Blake-Zisserman. However, in case of 70% outliers, the later is much better.

Overall, the simulation experiments clearly show that our relative pose selection criterion greatly improves calibration results in case of some camera pairs with additional outliers as well as additional noise. The smoothed information measure in combination with the Blake-Zisserman distribution is the best choice.

4.2 Real Data

The experiment on real data consists of two AVT Marlin monochrome cameras and six AVT Pike color cameras observing a scene, as depicted is Figure 5. We estimate the intrinsic camera parameters using Zhang's [16] calibration pattern based method. To be able to evaluate our calibration results, we use Zhang's method also to compute a "ground truth" for the extrinsic calibration. Note that this "ground truth" is *not* free of errors, but still provides a reasonable comparison. As input for our calibration, we extract 100 point correspondences in each image pair using SIFT [8].

Figure 6 contains the results for the Blake-Zisserman distribution. Our selection criterion using the smoothed information measure or the entropy measure provides greatly improved calibration results compared to the BFS traversal. This demonstrates that our method is actually beneficial in practical situations, where the quality of the correspondences between camera pairs cannot be strictly devided into "good" and "bad". The Cauchy distribution did not produce reasonable results at all.

5 Conclusions

In the context of calibrating a multi camera system via pairwise relative poses, we presented theoretically sound local and global uncertainty measures for relative poses: the purely local information measure, the purely global entropy measure and also the intermediate smoothed information measure. We formulated the selection of relative poses as a graph based discrete optimization problem, for which we presented an efficient algorithm. In the experiments, we showed that our selection criterion is able to greatly improve calibration results by avoiding bad relative poses caused by an increased outlier portion or increased noise. It turned out that the smoothed information measure, which takes local as well as global uncertainty into consideration, gives the best results. Furthermore, we showed that the Blake-Zisserman distribution is much better suited than the Cauchy distribution, especially in case of many outliers. The experiments on real data confirmed these findings.

References

[1] P. Baker and Y. Aloimonos. Complete Calibration of a Multi-camera Network. In *Proceedings of the IEEE Workshop on Omnidirectional Vision (OMNIVIS)*, pages 134–144, 2000.

[2] J. P. Barreto and K. Daniilidis. Wide Area Multiple Camera Calibration and Estimation of Radial Distortion. In *Proceedings of the IEEE Workshop on Omnidirectional Vision (OMNIVIS)*, 2004.

[3] M. Brückner, F. Bajramovic, and J. Denzler. Experimental Evaluation of Relative Pose Estimation Algorithms. In *Proc. of the Third International Conf. on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 431–438, 2008.

[4] X. Chen, J. Davis, and P. Slusallek. Wide Area Camera Calibration Using Virtual Calibration Objects. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 2520–2527, 2000.

[5] C. Engels and D. Nistér. Global uncertainty in epipolar geometry via fully and partially data-driven sampling. In *ISPRS Workshop BenCOS: Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images*, pages 17–22, 2005.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003.

[7] I. Kitahara, H. Saito, S. Akimichi, T. Ono, Y. Ohta, and T. Kanade. Large-scale Virtualized Reality. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Technical Sketches*, 2001.

[8] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

[9] D. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[10] W. E. Mantzel, H. Choi, and R. G. Baraniuk. Distributed Camera Network Localization. In *Proceedings of the 38th Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1381–1386, 2004.

[11] D. Martinec and T. Pajdla. 3D Reconstruction by Gluing Pair-Wise Euclidean Reconstructions, or "How to Achieve a Good Reconstruction from Bad Images". In *3DPVT '06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pages 25–32, 2006.

[12] D. Martinec and T. Pajdla. Robust Rotation and Translation Estimation in Multiview Reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.

[13] H. Stewénius, C. Engels, and D. Nistér. Recent Developments on Direct Relative Orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294, 2006.

[14] T. Svoboda, D. Martinec, and T. Pajdla. A Convenient Multi-Camera Self-Calibration for Virtual Environments. *PRESENCE: Teleoperators and Virtual Environments*, 14(4):407–422, 2005.

[15] J. Vergés-Llahí, D. Moldovan, and T. Wada. A new reliability measure for essential matrices suitable in multiple view calibration. In *Proc. of the Third Int. Conf. on Comp. Vision Theory and Applications (VISAPP)*, volume 1, pages 114–121, 2008.

[16] Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.