# Robust Geometric Self-Calibration of Generic Multi-Projector Camera Systems

Simon Willi* Anselm Grundhöfer†

Disney Research

**ABSTRACT**

Calibration of multi-projector-camera systems (MPCS) is a cumbersome and time-consuming process. It is of great importance to have robust, fast and accurate calibration procedures at hand for a wide variety of practical applications. We propose a fully automated self-calibration method for arbitrarily complex MPCS. It enables reliable and accurate intrinsic and extrinsic calibration without any human parameter tuning. We evaluated the proposed methods using more than ten multi-projection datasets ranging from a toy castle set up consisting of three cameras and one projector up to a half dome display system with more than 30 devices. Comparisons to reference calibrations, which were generated using the standard checkerboard calibration approach [44], show the reliability of our proposed pipeline, while a ground truth evaluation also shows that the resulting reconstructed point cloud accurately matches the shape of the reference geometry. Besides being fully automatic without the necessity of parameter fine tuning, the proposed method also significantly reduces the installation time of MPCS compared to checkerboard-based methods and makes it more suitable for real-world applications.

**Keywords:** Projector-camera systems, Calibration and registration of sensing systems, Display hardware, including 3D, stereoscopic and multi-user Entertainment, broadcast

## 1 INTRODUCTION AND MOTIVATION

Accurately calibrating cameras is a core requirement for numerous computer vision tasks. A huge amount of research has been carried out within this field to solve this task either manually or automatically up to full self-calibrations allowing Rome to be reconstructed within a day [1]. All these methods vary in their complexity, prerequisites, and accuracy.

Projector-camera systems have been used for several decades within different application fields, such as 3D scanning applications, light transport decomposition [30], spatial augmented reality (also known as projection mapping) [27], interactive installations, and adaptive and dynamic color correction tools. All of these setups require some kind of global device registration, which is achieved by using the cameras to capture projector patterns and estimate the needed information for calibrating the projectors. This task is mostly carried out in a preprocessing step. Since this is time consuming and error prone, other methods propose to apply self-calibration algorithms to achieve this goal. Since, however, the number of camera views is usually quite low and contains significant perspective variations, not all methods (such as the ones optimized for structure from motion) can work reliably. Although, recently, researchers have proposed several MPCS self-calibration methods, they are all limited to specific setups or require additional information, such as the geometry to be known or initial estimates for the camera intrinsics.

The main goals of our work are to overcome these limitations and develop a generic, reliable, and outlier insensitive self-calibration method which is sufficiently flexible to handle all kinds of MPCS, fast to compute, and does not require any initial guesses or manual parameter tuning.

## 2 BACKGROUND AND RELATED WORK

Projector camera systems, also called procams, are combined input and output devices being used, for example, for surface scanning or augmentation tasks. They contain cameras which observe the projection onto the surface. Depending on the application purpose, this information can be used to generate an accurate geometrical calibration of the projector with respect to the real world. Sample applications are presented, for example, in [3] and [27].

### 2.1 Geometric Calibration of Projector-Camera Systems

Most methods to calibrate projectors usually start with one or multiple cameras, which are either pre-calibrated or uncalibrated during the process. Calibrating the intrinsics of cameras can be carried out in various ways, the most widely used ones involve multiple captured images of a planar marker board of unknown orientation, often with a checkerboard pattern, to estimate the focal length, principal point, and specific amount of parameters to model the lens distortion. The most commonly used method to apply this calibration is presented by Zhang et al [44], which is also widely used as the baseline to compare other calibration methods. Since the accuracy of such methods strongly depends on the number of samples and marker orientations within the various captured images, in [33], efforts were made to actively assist users in selecting the most useful poses to generate an accurate calibration result: The authors propose an iterative method to estimate the most suitable marker orientation for the next capture from the current calibration results, which could be shown to improve the calibration accuracy. The same checkerboard method can also be used to calibrate the relative orientation of the cameras, i.e. the extrinsic properties, by presenting the same patterns in different camera views. Having at least two calibrated cameras, a projector calibration can be carried out using structured light patterns to generate correspondences between all cameras and projectors. Since the cameras are already calibrated, the correspondences can be used to triangulate a point cloud of the surface and then use this information to register the projectors to the surface.

To avoid the requirement of using multiple calibrated cameras for projector calibration, methods were proposed to calibrate the projector via planar surfaces in an equivalent fashion to the aforementioned camera calibration, but this time, by treating the projector as an inverse of a camera and using structured light patterns to estimate with a single camera, where each projector pixel is seen on a planar surface [31, 8, 29, 24]. Recently, this process was simplified by using self-identifying projected blob patterns, which can also be robustly detected when projected onto planes which are

*e-mail:simon@disneyresearch.com

†e-mail:anselm@disneyresearch.com

placed significantly out of the focus plane of the projector [43]. Related plane-based methods are presented in [9], which also nicely summarizes further-related methods and differences between them.

The RoomAlive system presented by Jones et al [21] uses multiple Kinect depth cameras to register projectors to a static geometry. Although the system can be used to set up living-room-scale projection based augmentations, the noise characteristics, sensing range and limited resolution of the depth sensors do not provide a high-quality, pixel-accurate calibration in very large rooms (up to domes).

If the geometry of the projection surface is known, manual correspondences can also be generated without using a camera [7]. However, besides the fact that this is often not the case, this process is cumbersome and error prone.

### 2.2 Self-Calibration Methods

One of the first methods to fully automatically calibrate a generic projector-camera pair without using a planar surface has been proposed by Yamazaki et al [42]. They propose an algorithm based on the decomposition of a radial fundamental matrix into intrinsic and extrinsic parameters, which requires a close-to-pixel accurate prior for the principal point. This is hard to achieve in real-world situations, where it is usually not located close to the center of the image plane but shifted on the y-axis due to the lens shift optics. In [34], Sajadi and colleagues present a system which enables the calibration of multiple cameras and projectors, assuming that the cameras all share the same focal length and no distortion parameters, which can be hard to achieve depending on the used lenses.

Garcia et al [12] proposed a method to calibrate a specific MPCS in which sensors face each other and share a common viewpoint using translucent planar sheets placed at a series of varying orientations to generate planar pixel correspondences between all devices. Using this information, the standard method presented by Zhang [44], with an additional sparse bundle adjustment (SBA) step [40], is used to calibrate the devices. Although the approach is able to accurately calibrate multi-projector setups, it is limited for a specific configuration and, thus, cannot handle the desired variety of complex setups. Recently, a method was proposed by Garrido-Jurado et al [13], which offers a flexible self-calibration method. Although they focus on the same goal as our work, their approach has several limitations. The most important one is the fact that the authors assume that the intrinsics of the devices are already known beforehand, which is quite often not the case, especially since zoom and focus are often readjusted for each particular setup. Because of that limitation, their strategy to insert new devices focuses solely on the number of available correspondences and how to optimize the device integration strategy using a mixed integer linear programming approach. Although their method showed convincing results for a specific setup, the requirements of having the intrinsics pre-calibrated, no direct outlier treatment, and a relatively simple integration strategy when compared to, for example, the strategy proposed by [37] makes it less flexible for generic usage. Another recent method presented by Li et al [23] uses priors for the principal points as well as for the focal lengths. While the principal point can be roughly estimated to be in the center for cameras, this is usually not the case for projectors. In order to estimate them the authors propose a method that requires the zoom level of the projector to be changed. Not only that not all projectors do have different zoom levels but also changing the zoom has usually to be done manually it is impractical for bigger MPCS. Furthermore a rough estimate of object size and distance is required for the focal length priors. The semi-automated method presented by Fleischmann et al [11] uses projected vanishing points to estimate the internal and external calibration parameters. A physical projection surface with three mutually orthogonal planes as well as user guidance is required by this approach. Another semi-automated self-calibration method is proposed by Resch et al [32] which also accurately recovers the global scale. The latter is derived from the projection surface geometry which needs to be known beforehand. In contrast to their work, we are targeting a generic solution which can also handle situation in which the surface geometry is totally unknown.

Our approach focuses on a generic method to calibrate MPCS for complex projection mapping installations in which the used cameras, projectors, and optics can vary, not only from setup to setup but also within particular installations. Therefore, we focus on a self-calibration method that requires the least amount of initial assumptions and constraints.

### 2.3 Contribution

Although we are not the first researchers to tackle the problem of a full geometrical self-calibration of MPCS, we present the first method, which can be successfully applied to a wide variety of setups, without the need of further modifications, calibration prerequisites, and manual parameter tuning. The following points summarize our main contributions:

- Our proposed method enables to fully automatically calibrate arbitrary complex MPCS with a minimum number of two cameras and one projector without an upper limit.
- The method is able to automatically calculate the most robust sequence of device calibration.
- No initial intrinsic parameters need to be known beforehand.
- The system is able to efficiently make use of the knowledge of projector-to-camera pixel correspondences.
- Outliers are detected automatically and excluded effectively during the calibration process
- No manual parameter adjustment is required.
- The method is straightforward to implement and able to calibrate large datasets from scratch within several minutes.

We evaluated the method on a range of real-world datasets of varying complexities and sizes which could all be successfully calibrated using the proposed method without any manual parameter adjustment. In the next section we will describe the individual steps of our proposed method, followed by an evaluation and comparison with reference calibrations.

## 3 Robust Geometric Self-Calibration of Generic Multi-Projector Camera Systems

The main goal of our work is the generation of a robust, reliable, and flexible MPCS calibration. This should be easy to carry out, fast to process, and sufficiently insensitive to work with noisy and partially faulty data, which is quite often the case during complex real-world installations. Furthermore, the algorithm should be insensitive to the hardware used which, currently, might range from XGA up to 4K projectors, as well as VGA machine vision cameras or DSLRs containing sensors with dozens of megapixels. Our method is currently focused on devices approximating a perspective pinhole projection.

Therefore, each of the individual steps are focused on these specific requirements and will be explained in detail in the following sections, starting with the preprocessing, to generate accurate pixel correspondences, followed by the significantly important step of the initial pair selection and the outlier insensitive global system calibration.

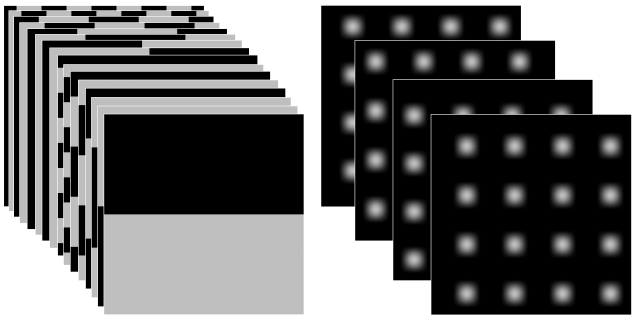


Figure 1: A simplified structured light pattern sequence to generate sub-pixel accurate correspondences between the devices. Complementary Gray codes are used to reliably generate coarse references. The additionally projected blob patters are then used to refine them down to sub-pixel accuracy. Depending on the application requirements, all, or only a subset of all projector pixels, are used for correspondence generation.

### 3.1 Preprocessing: Pixel Correspondence Generation

Most pinhole device calibration algorithms usually require some kind of correspondence information between either world features and image plane or between the image planes of different devices to calibrate. While this is a non-trivial task in standard multi-view reconstruction methods and requires feature detectors and matchers, such as, for example, SIFT[25] or SURF[2], within MPCS, the correspondences can be generated reliably using projected patterns which uniquely encode information about the according projector pixels. To generate a large amount of sub-pixel accurate correspondences between the different devices, we project such structured light patterns and capture them with the cameras. A large variety of patterns with different pros and cons exist, and an in-depth evaluation of the existing strategies is out of the scope of this paper (the interested reader is referred to [35] and [36] for an overview). We developed a method which has been adapted from Gray code [20] patterns in combination with the sub-pixel accurate line shift approach presented by [17]. Since the latter has accuracy issues when carried out on complex shapes and tilted cameras and we are focusing on a reliable method for generic diffuse surfaces, we developed a robust structured light method using complementary Gray code patterns [6], plus several dense blob patterns for further sub-pixel accurate refinement (cf. Figure 1, which presents some simplified sample patterns). After projecting and capturing these patterns, they are used to generate a series of precise correspondence maps between each individual projector and camera. To speed up the process, only a subset of all projector pixels are used for correspondence generation. Usually, several thousand pixels are used, but the method can also be applied to generate mappings for all projector pixels, if desired, for example, for surface reconstruction. However, this enormous amount of correspondences is not required for calibration. Depending on the camera capture rate, the projection of these patterns usually takes several seconds up to minutes per projector. All cameras can obviously capture one projection in parallel. Currently, the processing takes approximately $20s$ per device pair, using a straightforward CPU implementation, but might vary depending on the number of blobs shifts and device resolution.

Most structured light projection methods are intended to operate on mostly diffuse surfaces [36], so is our method. Non-diffuse surfaces can generate illumination effects such as indirect reflections, scattered light, caustics, refractions, and highlights. These effects might lead to non-unique mappings of projector pixels to camera coordinates. Another cause for false mappings can be uncontrollable, dynamic light sources during the acquisition process. In such cases, influenced areas have to be masked out before processing. However, any static illumination patterns such as, for example, exit sign lights are fully automatically removed by capturing a minimum and maximum intensity image of each projector and discarding all areas without any significant change from the correspondence estimation. To our knowledge the mentioned limitations are common in all structured light methods and not specific to our approach.

### 3.2 Correspondence Analysis and Initial Pair Selection

Having computed all projector-to-camera correspondences, they are transformed into a series of camera-to-camera correspondences $C2C_{ij}; i, j \in [0, k-1]$ for all $k$ cameras. This results in

$$m = \sum_{i=1}^{k-1}(k-i) \tag{1}$$

camera-to-camera maps. For each projector pixel which has been observed by both cameras via the structured light patterns, the maps store the according floating point image plane coordinates on the $i$-th and $j$-th camera, as well as the index to the according projector. As mentioned in [18], [13], and [38], finding the optimal camera pair to initiate the self-calibration process is a highly critical step to achieve an accurate result. If this initial calibration fails due to too many outliers or a nearly degenerate configuration, the whole calibration process is likely to fail. Although several schemes to determine the optimal pairing for multi-camera setups have been presented in the literature, MPCS incorporates further constraints since it is known that the projector pixels are arranged in a regular two-dimensional grid. Using this information, we developed a reliable voting scheme to rank the different $C2C$ correspondence maps with respect to their usability to initiate the self-calibration process. Each individual step is described in the following and an overview of the initial pair selection scheme is provided in Figure 3.

Unsuitable Pair Removal First, all $C2C$s are analyzed for their number of pixel correspondences and a threshold $t_{corr} = 100$ is defined. Any $C2C$ containing less than $t_{corr}$ correspondences will be removed from the initial pair selection process to increase stability.

Outlier Removal For all remaining $C2C$'s, the Fundamental matrix $F_{ji}$ between both cameras is estimated with a robust algorithm [16], using RANSAC [10] to make the procedure insensitive to outliers. During this process, the outlier threshold is set to an aggressive value of $t_F = 0.00004 * min(max(w_i, h_i), max(w_j, h_j))$, where $w_i, h_i$ is the width and height of the i-th image. The resulting $t_F$ is only a fraction of the one proposed by [37] to ensure that all outliers were excluded. If the number of detected inliers falls below $t_{corr}$, the according $C2C$ gets removed from the voting process. Otherwise, we estimate the focal lengths of both devices using Bougnoux's formula [5]:

$$\begin{aligned} f_i^2 &= -\frac{p_j^T[e_j] \times \hat{I}_3 F_{ji} p_i p_j^T F_{ij} p_j}{p_j^T[e_j] \times \hat{I}_3 F_{ji} \hat{I}_3 F_{ij} p_j} \\ f_j^2 &= -\frac{p_i^T[e_i] \times \hat{I}_3 F_{ij} p_j p_i^T F_{ji} p_i}{p_i^T[e_i] \times \hat{I}_3 F_{ij} \hat{I}_3 F_{ji} p_i} \end{aligned} \tag{2}$$

where it is assumed that the principal points in homogeneous coordinates $p_i$ and $p_j$ of the cameras are located at the respective image centers. $[e_i]$ and $[e_j]$ are skew-symmetric matrices of the left and right null vectors of $F_{ij}$, and $\hat{I}_3$ defined as:

$$\hat{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3}$$

This method is very sensitive to already small errors in the estimation of $F$ and can potentially generate negative values, which does not allow to draw the required square root to get the focal length. If this is the case, the according camera pair is also excluded from the voting process. In case both focal lengths are positive than $f_i$, $f_j$, and the according Fundamental matrix $F_{ij}$ are stored for later processing, and all detected outliers within $C2C_{ij}$ are removed.

Overlap Computation For all remaining ones, the normalized overlaps on the image planes are computed: For all remaining entries in $C2C_{ij}$, we find the two axis-aligned bounding boxes that contain all image plane correspondences to the cameras $i$ and $j$. To avoid a device-dependent bias, the width $bb_w$ and height $bb_h$ of the bounding boxes are normalized to generate a resolution-independent measure. For both cameras, we calculate the normalized area of their bounding box $A = bb_w * bb_h$, and multiply the two normalized areas $A_{i\leftrightarrow j} = A_{i\rightarrow j} * A_{j\rightarrow i}$. If $A_{i\leftrightarrow j}$ is less than $t_A = 0.01$ which means $1\%$ of the multiplied normalized image planes, the pair gets discarded as well since the small area might lead to a poor calibration accuracy.

Gradient Difference Estimation Having removed all potentially unsuitable $C2C$'s, the remaining ones are analyzed to estimate whether they have a potentially wide baseline and are observing a sufficiently varying surface to avoid a degenerate configuration. To estimate how different the cameras $i$ and $j$ are observing the unknown projection surface, all $C2C$s are analyzed for their 2D gradient difference on each projector image plane. As already mentioned, these maps store indices to the according projector for each correspondence. From this information, the individual projector correspondences are separated and used to compute a per-projector gradient map, which stores the normalized local gradient change of the camera correspondences on the projector's image plane: for each entry in the $C2C$, the according normalized pixel correspondence to camera $i$ and $j$ is selected, and its difference to the closest next correspondence on the projector image plane is computed. This value then is normalized by dividing it by the normalized $L2$ distance on the projector pixels. This is done individually for all correspondences per projector. Since we know that the projector projects pixels in a regular manner, the variance within these gradient changes are a measure of the projection surface variation. Since we have two gradient maps $\nabla(i)_{xy}$ and $\nabla(j)_{xy}$, where $x$ and $y$ are all projector pixels storing correspondences, we can compare them to estimate whether both cameras observe the surface from different directions. Therefore, the two gradient maps are used to generate a series of absolute gradient differences by subtracting both values:

$$\Gamma(ij)_{xy} = |\nabla(i)_{xy} - \nabla(j)_{xy}| \tag{4}$$

From these maps the mean $\mu(\Gamma(ij))$, and standard deviation $\sigma(\Gamma(ij))$ are computed to indicate whether both devices $i$ and $j$ are well suited for an initial calibration step: The higher both values, the more likely the cameras observe the surface from significantly different observation angles and orientations. Figure 2 gives a schematic explanation for this process.

Voting With all this information, votes are computed for each remaining $C2C$ to select the best initial pair:

$$V_{(ij)} = A_{i\leftrightarrow j} * (\mu(\Gamma(ij)) + \sigma(\Gamma(ij))), \tag{5}$$

weighting the pairs with the normalized overlap area multiplied by the sum of the mean and standard deviation of the gradient differences as most suitable for calibration. From all $Vs$, the highest ranked correspondence pair is chosen for an initial calibration and point cloud reconstruction.

### 3.3 Initial Pair Calibration

Having chosen the initial camera pair, the self-calibration process is initiated by estimating a local intrinsic and extrinsic calibration for these two devices. In a first step, we therefore use the already estimated $F_{ij}$, which we can use to estimate $f_i$ and $f_j$, as described in the last section. To account for potential matrix degradation due to lens imaging imperfections, distortion parameters are estimated and optimized together with $F_{ij}$, as well as $p_i$ and $p_j$, in a constrained non-linear minimization step: Assuming that the generated initial guesses for $f_i$ and $f_j$ are relatively accurate, we constrain the optimized focal lengths to not deviate more than 50% from the initial guess during the optimization, which successfully limits the risk to optimize for too large distortion parameters.

For the optimized $F_{ij}$, $p_i$, and $p_j$, we again estimate the focal lengths of both devices using Bougnoux's formula [5] (cf. Equation 2) and assemble the two calibration matrices $C_i$ and $C_j$. Having computed the latter, the essential matrix $E$ is computed by:

$$E_{ij} = C_i^T * F_{ij} * C_j. \tag{6}$$

For a more reliable estimation, $E$ is constrained since we know that $SVD(E) = U * W * V^T$, as proposed by [18]. With

$$W_c = \hat{I}_3 \tag{7}$$

as defined in Equation 3 and computed using:

$$E_c = U * W_c * V^T. \tag{8}$$

For $E_c$, we can finally estimate the relative extrinsic values for rotation $R$ and translation $t$ between the two cameras $i$ and $j$, as also described in [18].

For this step, a pair of intrinsic and extrinsic calibration data is computed, and a point cloud for all point correspondences within $C2C_{ij}$ is calculated by triangulation using iteratively weighted least squares as proposed in [19].

Outlier Removal Having generated the 3D point cloud, an iterative refinement procedure is started similar to the one proposed in [38] but adaptive, as well as resolution independent. A threshold to classify outliers is generated depending on the re-projection errors for the $3D \leftrightarrow 2D$ correspondences to both cameras. These errors are transformed into a resolution-independent value by setting the diagonal to 1000 pixel ($diag == 1000$) such that cameras having significantly varying resolutions do not lead to a biased re-projection error estimation. For each $3D$ point, the largest (normalized) reprojection error of the devices $i$ and $j$ is stored in a list from which, finally, the mean and standard deviation are computed to estimate an adaptive outlier threshold, defined as $t_{adapt} = mean + 2 * stddev$.

The calculated re-projection errors are clamped to $err = min(max(err, 1.0), 100.0)$, and all $3D$ points with $err > t_{adapt}$ are removed. The remaining points, as well as the intrinsic and extrinsic camera parameters of both devices, are further optimized in a sparse bundle adjustment (SBA) step [41], and the outlier removal strategy is applied again. These two steps, outlier removal followed by an SBA optimization, are repeated as long as the change of the average reprojection error is larger then $10\%$ compared to the last iteration and there has been at least one point removed during the outlier removal step.

### 3.4 Consecutive Device Integration

As soon as the first reconstruction is carried out and optimized, the remaining devices can be directly integrated iteratively by using the now existing $3D \leftrightarrow 2D$ correspondences. Therefore, the remaining cameras are integrated in an order depending on the amount of existing correspondences by carrying out the following steps: The

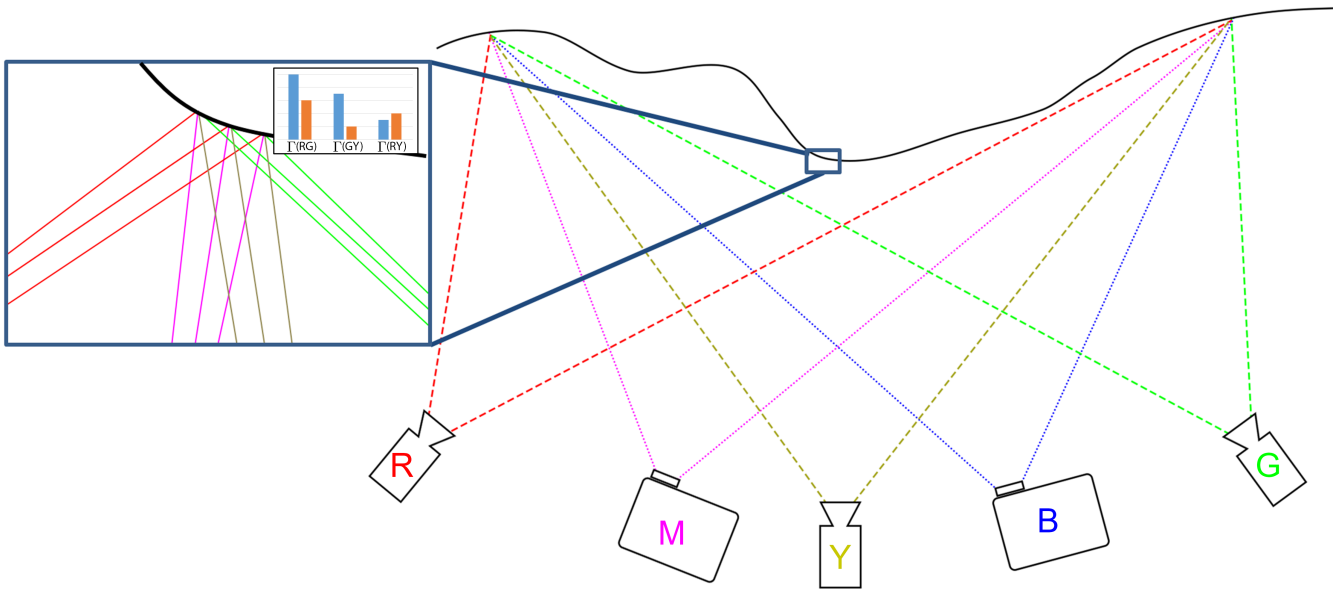



Figure 2: Initial pair selection: The regular projector pixel pattern is used to estimate local pixel difference gradients for each captured camera. This information is then used to estimate the camera pair in which the mean gradient difference plus its variance is maximized. In the zoom-in, $\Gamma(RG)$ is larger and more diverse than $\Gamma(RY)$ and $\Gamma(GY)$ for the two gradients (light blue & orange) between the three highlighted correspondences to projector M. This measure is calculated for all pixels and, in combination with the overall $C2C$ camera sensor area coverage, the most well suited camera pair for initiating the calibration process is chosen.

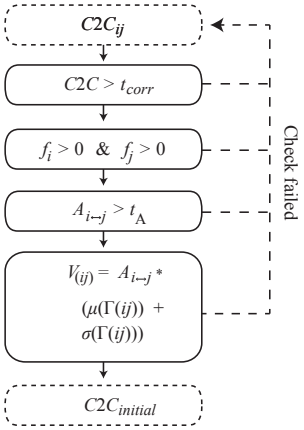



Figure 3: Flowchart for the different selection criteria for the initial C2C pair. Only device pairs that fulfill all criteria (in solid boxes) are candidates for being calibrated as first devices. If one of the checks fails, the next C2C pair is evaluated (dashed arrows) .

direct linear transformation (DLT) method [39] is used to generate a first initial guess of the new device's calibration data. This is further refined, and distortion parameters are estimated in an additional non-linear optimization step. The point cloud is now regenerated and extended by also triangulating the newly added correspondences of camera $l$, which had not been available within the point cloud before, i.e. all the correspondences which are not shared by cameras $i$, $j$, and $l$, but only exist in $C2C_{il}$ and $C2C_{jl}$. Then, one iteration of the same outlier removal strategy is applied to this dataset.

Next, the new camera, as well as the 3D point locations, are further refined by applying SBA by fixing the calibration data of all other cameras and another outlier removal step is applied again, as long as the average reprojection error has changed more than 10% and at least one outlier has been removed. After that step, a full SBA optimization is applied, optimizing all points and cameras, and outlier removal is carried out again in the same fashion, as mentioned above. In order to accelerate the relatively expensive SBA calls, for each device within the SBA call, a random subset of a maximum of 500 inlier points per device is used during these operations.

At this point, the newly added camera is fully integrated. However, since many correspondences were potentially removed during the outlier analysis, all initial point correspondences are now triangulated again, and only one single outlier removal step is applied afterwards before proceeding with the next camera device. These steps are repeated until all cameras are calibrated.

Having calibrated all cameras, the projectors are subsequently integrated into the reconstructed point cloud following the same strategy as before until no device is left.

Finally, all devices are successfully geometrically registered into one global coordinate frame, and the system is ready to be used for projection. The overall process flow is illustrated in Figure 4. The generation of consistent, blended, and color adjusted content is out of the scope of this work, thus, the interested reader is referred to [26] and [3].

## 4 Evaluation

Since the main goal of this work is the reliable calibration of arbitrarily arranged, intrinsically and extrinsically uncalibrated MPCS, we evaluated the proposed algorithm on a variety of datasets with different complexities (four to 33 devices) and volume sizes (surface width from $0.5m$ up to $30m$). The projectors varied in their lenses, as well as resolutions, from 800x600 up to $4K$. The cameras were either machine vision or DSLR cameras with a variety of lenses ranging form extreme wide angle to normal field of view and resolutions between one and 24 MP. We also set up a system with a

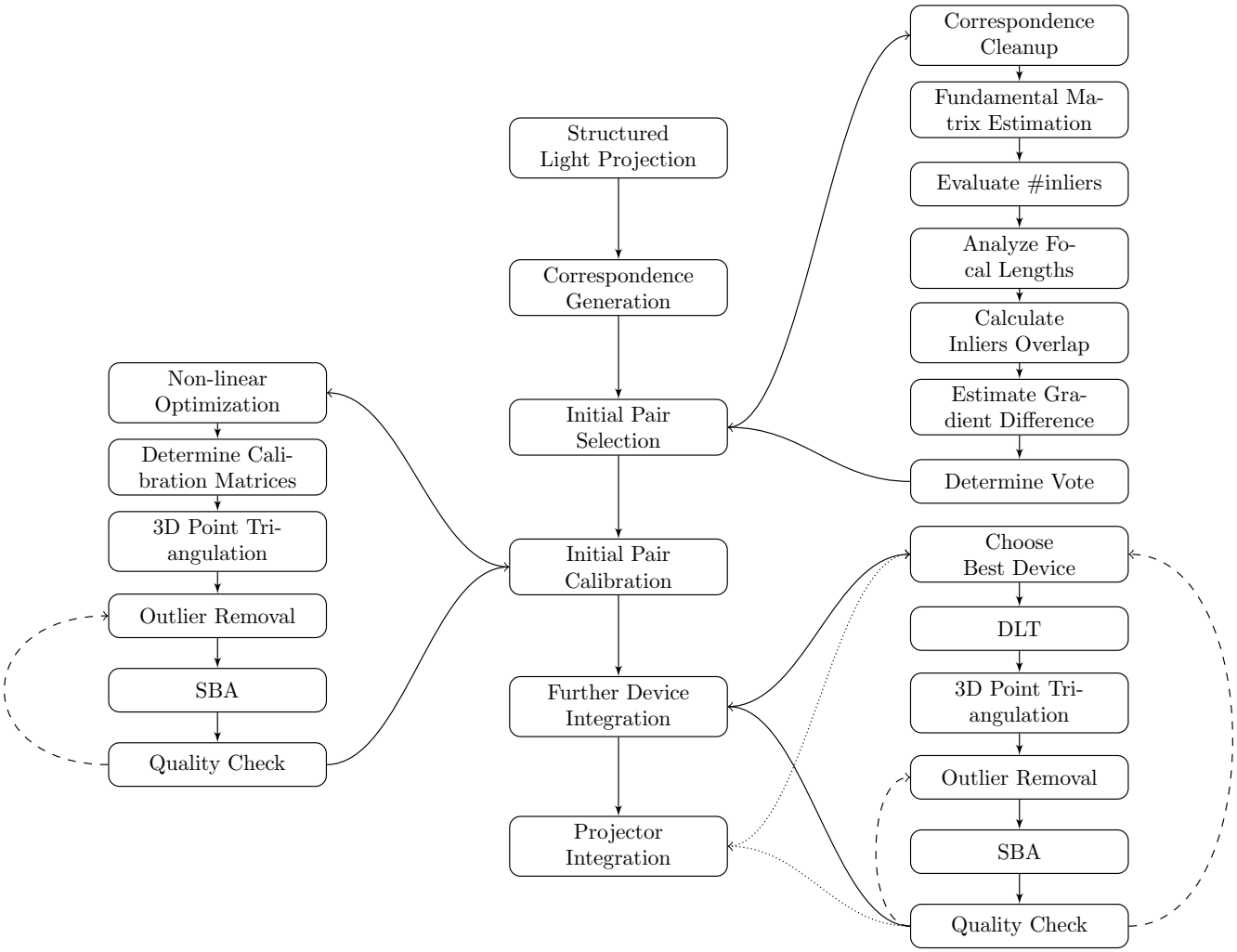



Figure 4: Simplified flow diagram of the proposed self-calibration algorithm. After the correspondence generation using structured light projection, the optimal initial pair selection is carried out considering outliers, spatial pixel correspondence distribution as well as their variations between the different devices. After that step, this pair is calibrated and an adaptive outlier removal step in combination with SBA is carried out. In the following all other cameras and projectors are integrated. Please refer to Section 3 for more details.

series of known issues, such as reflections, and complex geometry to evaluate how well the proposed method is able to calibrate such a setup. The algorithm has been implemented in C++ and all evaluations were carried out on a Intel® Xeon® CPU E5-1620 v4 @ 3.5GHz with 64 GB of RAM.

### 4.1 Universality Assessment

To evaluate the developed self-calibration method, we used datasets of 12 MPCS of a variety of setups ranging from small toy castle projections up to massive MPCS within half domes with dozens of devices. We applied our method to all of them using the default parameters as they were described in Section 3. It should be noted that no manual fine tuning was carried out at all. Details about the used datasets, as well as the results, are all summarized in Table 1 [1]. As it can be seen, our proposed method is able to fully automatically calibrate all of these datasets significantly faster than the time which would be required to carry out a high quality checkerboard calibration. Obviously, the processing time of our adaptive method increases with the number of devices used, but even the largest dataset, containing 33 devices, was calibrated accurately in less than 45 minutes. Please note that the resulting reprojection error, although normalized, has been sufficient to generate a seamless projection surface on all of them.

### 4.2 Comparison to Checkerboard Calibration

Dataset #9 and #12 were also compared to a checkerboard calibration carried out by an expert, using Zhang's checkerboard method for camera calibration plus additional DLT and SBA step. The resulting focal lengths of the devices are compared for #9 visually in Figure 5 and in table form for #12 (Table 2). As it can be seen, the differences are within a few percent, which makes both methods comparable in terms of accuracy.

### 4.3 Outlier Removal

One experimental setup (#11) generated several outliers during the structured light based correspondence generation due to surface re-

[1] Please note that due to confidentiality reasons the device placement, hardware details, and the reconstructed geometries cannot be presented for the datasets #1-#10.

Table 1: Summary of the 12 evaluated datasets. The application scenarios and system complexities varied a lot giving the wide range of surface structure, size and number of devices involved. Nevertheless, the proposed algorithm was capable to accurately calibrate all of them, as can be seen by the resulting average pixel re-projection errors (* To become device-independent, these values were all normalized by rescaling the dimension of the image plane diagonal to 1000, as described in Section 3). The # of 3D points lists the number of final reconstructed and outlier removed points at the end of the method

| | # cams | # projs | surface shape | processing time | reproj. err.* | std. dev. | # 3D points |
|---|---|---|---|---|---|---|---|
| #1 | 8 | 5 | Tube | 3m 35s | 0.0608 | 0.0526 | 34828 |
| #2 | 8 | 8 | Cave | 15m 36s | 0.6825 | 0.5741 | 126408 |
| #3 | 11 | 1 | Face | 5m 58s | 0.0700 | 0.0700 | 47681 |
| #4 | 3 | 1 | Toy castle | 0m 30s | 0.2322 | 0.2565 | 10886 |
| #5 | 9 | 24 | Half dome | 41m 40s | 0.0898 | 0.0619 | 196342 |
| #6 | 6 | 2 | Statue | 4m 39s | 0.1414 | 0.2137 | 66214 |
| #7 | 11 | 1 | Face | 6m 13s | 0.2540 | 0.6505 | 21289 |
| #8 | 8 | 6 | Rounded cube | 5m 11s | 0.1097 | 0.1016 | 38189 |
| #9 | 4 | 5 | Living room | 1m 33s | 0.1081 | 0.1249 | 17613 |
| #10 | 5 | 3 | Furniture pieces | 4m 20s | 0.4578 | 1.0686 | 98528 |
| #11 | 6 | 3 | Cardboard boxes | 3m 19s | 0.1353 | 0.2263 | 8406 |
| #12 | 4 | 2 | Wall corner (CAD) | 2m 6s | 0.0572 | 0.0917 | 52610 |

Table 2: Resulting focal lengths of the checkerboard and self-calibration of the ground truth data set.

| Device | Checkerboard | Self-Calibration | Difference |
|---|---|---|---|
| Allied | 6981.325 | 6984.001 | 2.676 |
| Canon 1 | 8011.988 | 8001.571 | 10.416 |
| Canon 2 | 2836.783 | 2837.006 | 0.223 |
| Ximea | 2272.990 | 2275.018 | 2.027 |
| BenQ Proj. 1 | 4154.256 | 4159.207 | 4.951 |
| BenQ Proj. 2 | 4171.227 | 4171.615 | 0.387 |

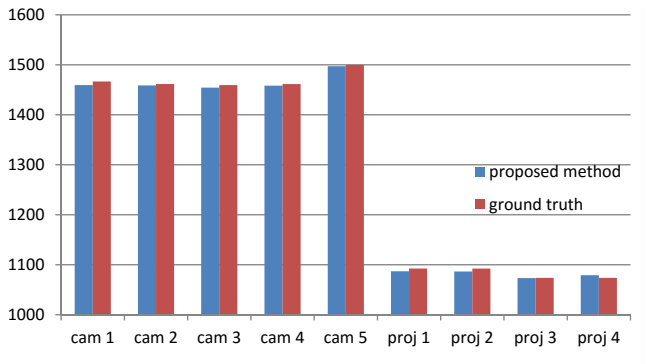



Figure 5: Focal length comparison of the devices of dataset #9, calibrated with the proposed self-calibration method, and a manual checkerboard calibration, carried out by an expert.

flections and scattering of light. An overview of the setup is given in Figure 6. This setup used DSLRs [2] with lenses, having focal lengths ranging from 24mm to 200mm, and used three projectors, one with short throw optics [3] and two with standard zoom lenses [4]. The outliers influenced the reconstruction accuracy, but since they are automatically estimated and removed during our reconstruction, they did not significantly influence the overall calibration process. As mentioned before, we are aware that the proposed structured light methods are only reliable using mainly diffuse surfaces; this setup has been used to demonstrate the robustness when used in those inappropriate situations.

### 4.4 Ground Truth Comparison

Another MPCS (#12) was generated using a computer generated geometrical model of a three-sided wall corner. This digital model was then used to manufacture a real-world setup with minimal deviations to enable a comparison of the self-calibration output to ground truth data (cf. Fig. 7). This heterogeneous system consists of four cameras [5] and two projectors [6]. The applied self-calibration required 2 min 6 sec to process and generate a calibration with re-projection errors as summarized in Table 3.

Since ground truth data was available for this setup, further evaluations were carried out. In order to compare the reconstruction as seen in 8 with the digital model (cf. Fig. 7a ) , fiducial 2D markers using the Aruco library [14, 15] were embedded in the projection surface (cf. Fig. 7b ). They were detected in the camera images and their spatial orientations with respect to the cameras were used to apply a global coordinate transformation. This transformation ensures that the extrinsic calibration resulting from the proposed self-calibration algorithm correlates to the coordinate system of the ground truth geometry definition. Since only a single consistent transformation should be applied to the individual extrinsics of the cameras and projectors, a Procrustes transformation [4] was applied which computed one consistent rotation and translation as well as a uniform scaling in all three dimensions for all estimated marker positions. To achieve a highly accurate transformation the following workflow was applied:

- Detect marker positions in all camera images
- Triangulate markers visible by at least 2 devices to generate

[2] 2xCanon EOS 600D, 3xCanon EOS 1100D, 1xCanon EOS 5D Mark II
[3] Benq MW843UST
[4] Mitsubishi MH2850U, Mitsubishi WL2650U
[5] 2xCanon EOS 1100D, 1xAllied Vision Manta G-504C, 1xXimea xiQ MQ013MG-E2
[6] 2xBenQ W1100

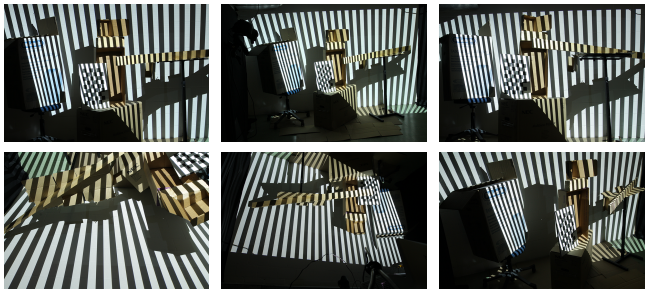
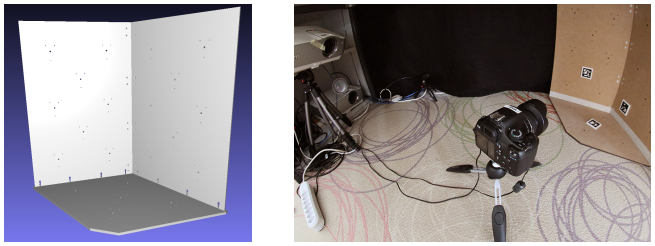


Figure 6: The six camera views for dataset #11. Even though the structure of the scene is rather complex including several reflections, our self-calibration procedure manages to successfully calibrate all devices within less then 4 minutes.



(a) Digital model (b) Setup

Figure 7: The ground truth model. Left: Rendering of the digital model, right: photograph of the real-world evaluation setup. The manufactured wall, both projectors and 2 of the 4 cameras are visible in the image.

3D point clouds

- Constrain 3D points positions ensuring equal marker sizes and planarity using a Levenberg-Marquardt optimization [28]
- Calculate Procrustes transformation between reference marker locations and the optimized reconstructions
- Apply single transformation to all cameras and projector orientations

Having registered the devices to the reference geometry, allows to also evaluate the reconstruction accuracy with respect to the known ground truth. Therefore, each vertex of the reconstructed point cloud was projected onto the three planes of the ground truth model and the distance to the closest one was evaluated. As it can be seen in Table 4, the deviation is relatively low (coordinate system scale 1unit = 1mm) with less than a millimeter average deviation to the ground truth which is approximately within the tolerances of

Table 3: Comparing checkerboard (CB) vs self-calibration (SC) reprojection errors. The average ∅ reprojection errors are normalized to a diagonal of 1000 pixels for the CB and SC approach. In parentheses ( ) the unnormalized values are given.

| Device | Resolution | ∅ error CB | ∅ error SC |
|---|---|---|---|
| Allied | 2452x2056 | 0.008 (0.026) | 0.008 (0.026) |
| Canon 1 | 4272x2848 | 0.044 (0.228) | 0.044 (0.228) |
| Canon 2 | 4272x2848 | 0.027 (0.140) | 0.027 (0.139) |
| Ximea | 1280x1024 | 0.020 (0.033) | 0.019 (0.031) |
| BenQ Proj. 1 | 1920x1080 | 0.042 (0.093) | 0.045 (0.099) |
| BenQ Proj. 2 | 1920x1080 | 0.065 (0.145) | 0.087 (0.192) |

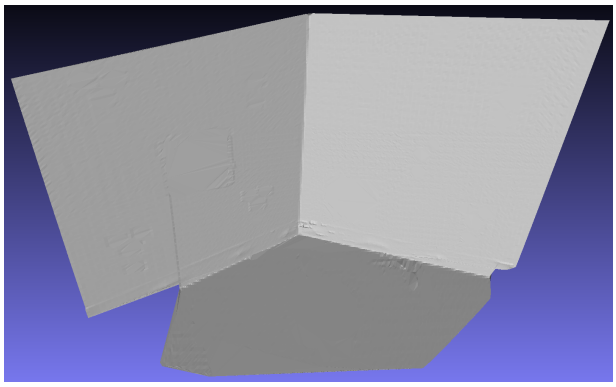


Figure 8: Reconstruction of dataset #12.

the manufacturing process. As it can be seen, the proposed method better approximates the geometry than the checkerboard calibration method. The resulting focal lengths can be compared in Table 2.

### 4.5 Scalability

Finally, a last evaluation is carried out to show the scalability of the proposed method. This MPCS consists of totally 8 cameras and 4 projectors. In Figure 9 we see how the normalized reprojection error changes depending on how many cameras are used to calibrate the system. We let the calibration procedure run with initially using only 2 cameras (cam6 and cam8), the remaining cameras not used at all. As the red bars in the corresponding devices show, the error is comparably large when using only two cameras. However, the method was still capable of successfully calibrate the 4 projectors and the 2 cameras. The error decreases rapidly when a 3th and 4th camera is used to calibrate the system. After that, adding more cameras might increase the errors again slightly since the system tries to find an optimal solution, which gets harder the more devices are involved. Figure 10 displays the relationship of required processing time, involved devices and number of vertices. Even with all 12 devices calibrated into the system, the calibration process took only 8 minutes.

## 5 Summary and Conclusions

In this paper, we proposed a robust and adaptive algorithm to enable a reliable self-calibration for arbitrarily complex MPCS. In contrast to previous research, it is not limited to specific setups (as long as at least two cameras and one projector is present) nor requires the geometry structure to be known in advance. It eliminates the need of expert knowledge to successfully calibrate the system and does not require any manual parameter tuning since it is self-adapting. Furthermore, it does not expect any initial focal length information and is able to reconstruct and calibrate even in situations where a significant amount of false correspondences are present. We demonstrated its flexibility by successfully calibrating a variety of different setups which go far beyond experimental and minimal lab setups. The comparison to reference calibrations shows the reliability of our proposed method. Finally, the algorithm is able to fully calibrate MPCS within a small time frame of less than a minute for systems consisting of a few devices but is also able to calibrate a complex system with more than 30 devices in less than 45 minutes.

Again, it should be noted that all of the evaluated MPCS varied not only in their complexity and physical size but also by the used camera and projection hardware. All the setups were entirely calibrated by the proposed adaptive method without applying any single manual parameter adjustment. Besides the lack of existence of any related generic self-calibration method not requiring initial

Table 4: Distance to the digital ground truth model. All units are in millimeters. As can be seen, the reconstruction of the proposed self-calibration approach closer resembles the ground truth data compared to a checkerboard based calibration.

| | avg | median | std deviation | 75th percentile | 99.9th percentile | max | min | # pts |
|---|---|---|---|---|---|---|---|---|
| Checkerboard | 1.4846 | 1.4913 | 1.1209 | 2.6340 | 4.6678 | 21.4142 | 2.83E-05 | 46074 |
| SelfCalibration | 0.9390 | 0.8481 | 0.6144 | 1.3726 | 2.1334 | 21.3255 | 8.68E-05 | 46074 |

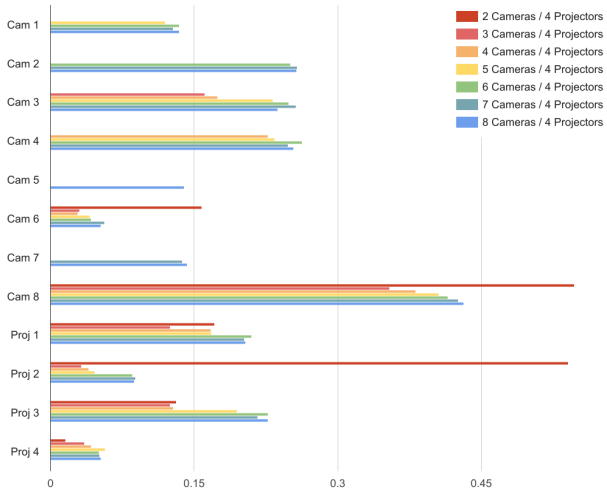



Figure 9: Calibration of the same setup with varying number of devices. The calibration was performed from only 2 cameras up to totally 8 cameras. The normalized reprojection errors are displayed along the x-axis. Although the reprojection erros were larger with only 2 cameras (red bars), the proposed self-calibration method was still capable to successfully calibrate all devices.

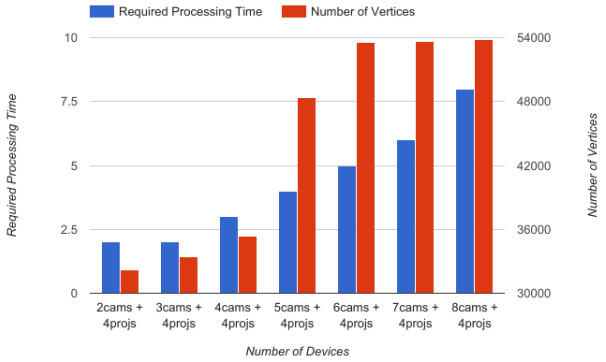



Figure 10: Required processing times and number of vertices in dependency of number of devices involved. Time units are in minutes.

guesses this is an evaluation that, to our knowledge, has not been carried out by any known related work so far.

Currently, we assume that all the devices use lenses with perspective projections. For the usage of fisheye lenses, the algorithm needs to be adapted to a distortion model which is as well suited for fisheye lenses as, for example, the one proposed by [22]. Implementing and evaluating this will be part of future research. Presumably, the weakest part of the current method is the requirement to have expert knowledge to reasonably choose the appropriate number and locations of camera views to successfully calibrate the system, which still requires a specific amount of expert knowledge to guarantee an accurate and reliable self-calibration. Overcoming this to facilitate the optimal camera placement is one future research direction.

## REFERENCES


[1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, Oct. 2011.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.

[3] O. Bimber and R. Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A. K. Peters, Ltd., Natick, MA, USA, 2005.

[4] I. Borg and P. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.

[5] S. Bougnoux. From Projective to Euclidean Space Under any Practical Situation, a Criticism of Self-Calibration. *ICCV*, pages 790–796, 1998.

[6] O. Choi, H. Lim, and S. C. Ahn. Robust binarization of gray-coded pattern images for smart projectors. In *2016 International Conference on Electronics, Information, and Communications (ICEIC)*, pages 1–4, Jan 2016.

[7] B. Close, D. B. McCulley, and B. H. Thomas. Arpipes: Aligning virtual models to their physical counterparts with spatial augmented reality. Adelaide, AU, 2010.

[8] I. Din, H. Anwar, I. Syed, H. Zafar, and L. Hasan. Projector calibration for pattern projection systems. *Journal of Applied Research and Technology*, 12(1):80 – 86, 2014.

[9] J. Draréni, S. Roy, and P. Sturm. Methods for geometrical video projector calibration. *Machine Vision and Applications*, 23(1):79–89, 2012.

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[11] O. Fleischmann and R. Koch. Fast projector-camera calibration for interactive projection mapping. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pages 3798–3803. IEEE, 2016.

[12] R. R. Garcia and A. Zakhor. Geometric calibration for a multi-camera-projector system. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*, pages 467–474, Jan 2013.

[13] S. Garrido-Jurado, R. Muoz-Salinas, F. Madrid-Cuevas, and M. Marn-Jimnez. Simultaneous reconstruction and calibration for multi-view structured light scanning. *Journal of Visual Communication and Image Representation*, 39:120 – 131, 2016.

[14] S. Garrido-Jurado, R. M. noz Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280 – 2292, 2014.

[15] S. Garrido-Jurado, R. M. noz Salinas, F. Madrid-Cuevas, and R. Medina-Carnicer. Generation of fiducial marker dictionaries using mixed integer linear programming. *Pattern Recognition*, 51:481 – 491, 2016.

[16] N. Gracias and J. Santos-Victor. Robust estimation of the fundamental matrix and stereo correspondences. In *Proc. of the International Symposium on Intelligent Robotic Systems*, Stockholm, Sweden, July 1997.

[17] J. Gühring. Dense 3-d surface acquisition by structured light using off-the-shelf components. In *Proc. Videometrics and Optical Methods for 3D Shape Measurement*, pages 220–231, 2001.

[18] R. Hartley and A. Zisserman. *Multiple view geometry in computer*

*vision*. Cambridge University Press, Cambridge, 2003. Choix de documents en appendice.

[19] R. I. Hartley and P. Sturm. *Triangulation*, pages 190–197. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995.

[20] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-d object recognition. In *International Conference on Pattern Recognition*, 1984.

[21] B. Jones, R. Sodhi, M. Murdock, R. Mehra, H. Benko, A. Wilson, E. Ofek, B. MacIntyre, N. Raghuvanshi, and L. Shapira. Roomalive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, pages 637–644, New York, NY, USA, 2014. ACM.

[22] J. Kannala and S. S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1335–1340, Aug 2006.

[23] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. Clausi, J. Zelek, and A. Wong. Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping. *IEEE Transactions on Computational Imaging*, 3(1):74–83, March 2017.

[24] T. Li, F. Hu, and Z. Geng. Geometric calibration of a camera-projector 3d imaging system. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, VRCAI '11, pages 187–194, New York, NY, USA, 2011. ACM.

[25] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[26] A. Majumder and M. S. Brown. *Practical Multi-Projector Display Design*. A K Peters, 2007.

[27] M. R. Mine, J. van Baar, A. Grundhfer, D. Rose, and B. Yang. Projection-based augmented reality in disney theme parks. *IEEE Computer*, 45(7):32–40, 2012.

[28] J. J. Moré. The Levenberg-Marquardt algorithm: Implementation and theory. In G. A. Watson, editor, *Numerical Analysis*, pages 105–116. Springer, Berlin, 1977.

[29] D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, 3DIMPVT '12, pages 464–471, Washington, DC, USA, 2012. IEEE Computer Society.

[30] M. O'Toole, J. Mather, and K. N. Kutulakos. 3d shape and indirect appearance by structured light transport. In *CVPR*, pages 3246–3253. IEEE Computer Society, 2014.

[31] J.-N. Ouellet, F. Rochette, and P. Hebert. Geometric calibration of a structured light system using circular control points. In *3D Data Processing, Visualization and Transmission*, pages 183–190, 2008.

[32] C. Resch, H. Naik, P. Keitler, S. Benkhardt, and G. Klinker. On-site semi-automatic calibration and registration of a projector-camera system using arbitrary objects with known geometry. *IEEE Transactions on Visualization and Computer Graphics*, 21(11):1211–1220, Nov 2015.

[33] A. Richardson, J. Strom, and E. Olson. AprilCal: Assisted and repeatable camera calibration. *IEEE International Conference on Intelligent Robots and Systems*, pages 1814–1821, 2013.

[34] B. Sajadi, M. A. Tehrani, M. Rahimzadeh, and A. Majumder. High-resolution lighting of 3d reliefs using a network of projectors and cameras. In *2015 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video, Lisbon, Portugal, July 8-10, 2015*, pages 1–4, 2015.

[35] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recogn.*, 43(8):2666–2680, Aug. 2010.

[36] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004.

[37] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: Exploring photo collections in 3d. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 835–846, New York, NY, USA, 2006. ACM.

[38] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from Internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, 2008.

[39] I. E. Sutherland. Three-dimensional data input by tablet. *SIGGRAPH Comput. Graph.*, 8(3):86–86, Sept. 1974.

[40] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ICCV '99, pages 298–372, London, UK, UK, 2000. Springer-Verlag.

[41] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment - a modern synthesis. In *Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*, ICCV '99, pages 298–372, London, UK, UK, 2000. Springer-Verlag.

[42] S. Yamazaki, M. Mochimaru, and T. Kanade. Simultaneous self-calibration of a projector and a camera using structured light. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 60–67, 2011.

[43] J. . M. G. Yang, L. ; Normand. Practical and precise projector-camera calibration. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Merida, Mexico*, 2016.

[44] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *International Conference on Computer Vision (ICCV)*, pages 666–673, 1999.