Decoupling Photometry and Geometry in Dense Variational Camera Calibration

Mathieu Aubry, Bastian Goldlücke, Kalin Kolev, Daniel Cremers
TU Munich, Germany

Abstract

We introduce a spatially dense variational approach to estimate the calibration of multiple cameras in the context of 3D reconstruction. We propose a relaxation scheme which allows to transform the original photometric error into a geometric one, thereby decoupling the problems of dense matching and camera calibration. In both quantitative and qualitative experiments, we demonstrate that the proposed decoupling scheme allows for robust and accurate estimation of camera parameters. In particular, the presented dense camera calibration formulation leads to substantial improvements both in the reconstructed 3D geometry and in the super-resolution texture estimation.

1. Introduction

1.1. Camera Calibration and Geometry Estimation: A Chicken-and-Egg Dilemma

The problem of multi-view 3D reconstruction is one of the most fundamental and extensively studied problems in computer vision with numerous applications beyond its domain. Following recent improvements in digital photography, it has undergone a revolution in recent years and is now competitive to the most reliable techniques for 3D modeling [14, 18]. At the core of each multi-view reconstruction pipeline is the calibration of the cameras, i. e. the estimation of position, orientation and intrinsic parameters for each camera.

In the last and the first half of the current decade, great efforts have been focused on automatic camera calibration based on image information alone. As a result, we now have a number of publicly available software packages by Klein et al. [9] and Snavely et al. [15] which allow to automatically determine the camera parameters from a collection of images. Yet, in the context of image-based modeling, the question of where each camera was located is obviously tightly intertwined with the estimation of geometry and texture. A highly accurate estimate of the object's geometry/texture – as for example generated by the recent super-resolution approach of Goldluecke and Cremers [5] – should help to further improve the estimation of camera parameters. Analogously, a geometry/texture modeling method could benefit from a more precise camera calibration. In other words, the problems of camera calibration and geometry/texture estimation are highly coupled. As a consequence, any progress in one of these fields opens up new ranges in the other.

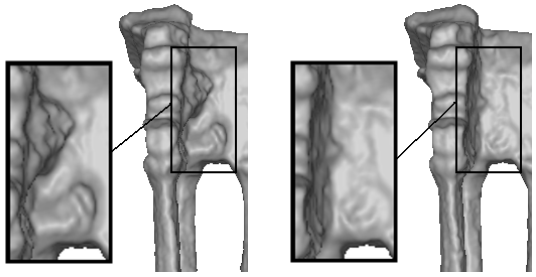


Figure 1. The proposed variational dense estimation of calibration parameters using a decoupling of photometry and geometry allows to drastically improve the estimation of 3D structure.

Up to date, it could be observed that while geometry and color are reconstructed in a *dense* manner, camera calibration methods typically rely on *sparse* feature correspondences. As the calibration problem is highly overdetermined (only 11 parameters are to be estimated for each camera in the full setting), a straightforward way to address it is to robustly pick a small subset of the provided information (*feature* points) for the estimation process. This naturally leads to the investigation of sparse feature-point based methods [8] which have become an established tool. Yet, the accuracy of the obtained parameters strongly depends on the precision of the underlying feature-point detector as well as the reliability of the matching procedure. Even though multiple heuristics have been proposed to tackle these tasks, like epipolar constraints and robust model fitting, the exploration of dense formulations to better understand the nature of the registration process deserves more attention (see figure 1).

1.2. Related Work

Sparse calibration, also referred to as structure-from-motion, has undergone exhaustive analysis. Research has

been conducted in various directions – ranging from the identification of salient image points [7, 12] and corresponding descriptors [12, 19] to the determination of minimal point sets needed [13] and their robust matching [3]. A key ingredient of a multi-camera calibration system is a global optimization step involving all camera parameters and estimated sparse 3D geometry, called *bundle adjustment* [20]. Usually, the calibration pipeline is split into multiple sequential stages: First, feature points and corresponding descriptors are estimated, then, an initial matching and 3D structure are established, and finally, a global refinement step is performed while filtering out mismatches and outliers. Thereby, the feature points are estimated only once at the beginning and held fixed during the entire calibration process so as to break the ill-posedness of the problem. Obviously, the accuracy of the obtained calibration parameters strongly depends on the precision of the underlying feature points and respective descriptors. Yet, in practice their precision is limited, since the computations are performed on a pixel basis without any knowledge of the observed geometry. In this work, we show how estimated dense 3D structure can be exploited to further improve a given calibration.

Recently, Furukawa and Ponce suggested a stereo-based approach for calibration refinement [4]. The method starts with initial calibration parameters and a sparse 3D point cloud representing the observed geometry and assigns an oriented patch to each point. The optimal calibration parameters and the precise localization of each 3D point in space are obtained by finding the local orientation giving rise to the most consistent patch distortion. Although this approach significantly improves upon classical sparse methods, it is still limited by the underlying local planarity assumption.

Multiple researchers have tried to use silhouettes to jointly estimate dense geometry and camera parameters. While the approaches of [22] and [2] require given object outlines in a binary form, the method of [23] is more general and proposes a unified framework for image segmentation and camera calibration. Although these techniques also have been shown to improve the calibration in certain cases, they are limited to specific objects and camera motion. For example, when panning around a spherical object, the observed silhouette may not change at all, which introduces a severe ambiguity in the optimization. In general, silhouette-based methods suffer from the fact that they do not exploit the whole available information, ignoring all color consistency between modeled and observed scene inside the object.

The refinement of camera parameters in spatially dense reconstruction methods has been further generalized in the variational approach of Unal et al. [21]. There, the authors suggest a generative model of image formation and subsequently estimate by gradient descent minimization both intrinsic parameters (focal length and skew) and extrinsic parameters (translation, rotation). However, the proposed formulation is based on particular assumptions regarding the radiance of the 3D scene – piecewise smoothness for the object and constancy for the background.

A closer analysis of this latter approach reveals that the camera parameters are estimated based on minimizing a *photometric* error (color difference between modeled and observed scene). Interestingly, this is fundamentally different from the *geometric* error (reprojection error) that most sparse state-of-the-art algorithms like bundle adjustment minimize. In contrast, the current work could be regarded as an effort for a dense formulation of bundle adjustment.

1.3. Contributions

In this paper, we revisit variational camera calibration in a spatially dense setting and propose a novel algorithm which is shown to provide more robust and accurate camera parameters. The proposed method improves over the work of Unal et al. [21] in several ways:

- The variational approach includes a super-resolution model of the object's texture, giving rise to more accurate camera parameters.
- We introduce a relaxation technique which allows to decouple the estimation of point correspondences and camera parameters. In this manner, we show that the original minimization problem can be solved by alternating dense correspondence estimation given by an optical-flow like algorithm with camera parameter estimation given by a continuous form of bundle adjustment.
- We experimentally demonstrate that the proposed variational camera calibration leads to substantial improvements both in the computed geometry and in the estimated texture.

2. Dense Calibration: Variational Formulation

2.1. Image Formation Model

Given a texture and a 3D model, the optimal camera parameters are those which minimize the reprojection error. We employ a super-resolution model of the reprojection error similar to the one formulated in [6]. Assume that we observe an object with known surface geometry Σ in n cameras, modeled by their projections $\pi_i : \mathbb{R}^3 \to \Omega$ mapping from 3D space into an image plane Ω. Let $\mathcal{I}_i : \Omega \to \mathbb{R}^3$, $i = 1, \dots, n$ denote the corresponding color images. Following the state-of-the-art super-resolution model [17], a real-world camera downsamples the input by integrating over the rays incoming in each sensor element. This process can be modeled by convolution with a kernel b derived from the properties of the camera [1].

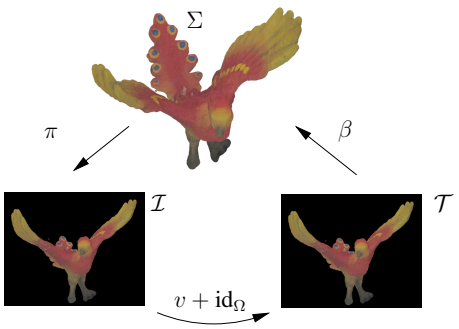


Figure 2. The image $\mathcal{T}$ is obtained by rendering the current 3D model and the respective texture map by means of the back-projection map β of each camera. We propose to use the optic flow v between $\mathcal{I}$ and $\mathcal{T}$ as a measure for the geometric reprojection error and minimize it with respect to the projection π.

Let $T : \Sigma \to \mathbb{R}^3$ be the estimated texture map of the surface extracted from the input images. Then, the reprojection error in terms of the texture and the unknown camera parameters π_k is given by

$$E(T, \pi) := \sum_{i=1}^{n} \int_{S_i} \|b * (T \circ \beta_i) - \mathcal{I}_i\| \, \mathrm{dx}. \qquad (1)$$

Here, the back-projection mappings $\beta_i : S_i \to \Sigma$ assign the visible point on the surface to each point in the silhouettes $S_i := \pi_i(\Sigma) \subset \Omega$, see figure 2. They satisfy $\beta_i \circ \pi_i = \mathrm{id}_\Sigma$ on the part of the surface visible in camera i, and so depend on π_i and Σ in a complicated way. Thereby, the convolution b is applied in the individual color channels separately. The term $T \circ \beta_i$ yields the observed intensity of the textured object in the view of the ith camera, which is transformed according to the super-resolution camera model and compared with the respective input image $\mathcal{I}_i$. The norm $\|.\|$ in (1) and in the following denotes the Euclidean norm.

While measuring the reprojection error in the individual images, the model in (1) differs from previously proposed ones on variational calibration [21] in two ways. Firstly, it relies on a super-resolution formulation which is expected to capture finer object texture details and thus leads to a more precise calibration. Secondly, we propose to exchange the L_2-norm used in [21, 6] with the L_1-norm which is theoretically more robust and which we found to give better results in practice.

2.2. Camera Reprojection Error

It was shown in [6] how to compute an accurate, super-resolved texture map by minimizing the energy (1). In this work, we focus on improving the camera calibration. That is, we assume the texture T as well as the 3D model Σ are pre-computed using an approximate initial camera calibration and minimize the energy (1) with respect to π in order to get a more accurate calibration. Of course, both the 3D model as well as the texture map can then iteratively be improved once the calibration becomes more accurate. For geometry reconstruction, we employ the algorithms proposed in [10] and [11]. While the energy optimized in their work to obtain the surface is different from the one above, it is still closely related to our variational approach.

Note that if T is kept fixed, each term of the sum in (1) is completely independent of the others, and can be minimized separately. For this reason, we will consider in the following only a single term of the sum and omit the dependence on the index i to simplify notations.

The derivatives of the back-projection β are very difficult to compute and depend on the 3D model whose accuracy is hard to predict. For this reason, we transform the energy (1) onto the surface and obtain

$$E_{\text{cam}}(\pi) = \int_{\beta(S)} \|\mathcal{I} \circ \pi - T\| \det(D\pi) \, \mathrm{ds} \qquad (2)$$

as the contribution of a single camera to the total reprojection error E. Thereby, $\det(D\pi)$ is the Jacobian of π which accounts for surface area foreshortening in terms of projection distortion [16]. Integration takes place over the back-projected silhouette $\beta(S) \subset \Sigma$, i.e. the part of the surface visible in the respective camera. We exploit the super-resolution model only for computing an accurate texture map, since it would be computationally prohibitive to optimize it with respect to the projections. Therefore, in (2) we set the kernel b to identity.

2.3. Direct Minimization via Gradient Descent

In order to minimize energy (2) with respect to π, we need a suitable parametrization of π by a set of parameters $(g_i)_{1 \le i \le m}$. In our implementation, we limit ourselves to the camera extrinsics, i.e. rotation and translation, giving rise to $m = 6$ degrees of freedom. Yet, a generalization to a more complex camera model is straightforward.

The simplest approach to minimize energy (2) is to perform gradient descent in the parameters g_i. In order to avoid an expensive recomputation of $\mathcal{I} \circ \pi$ in each step, one can switch to a Taylor expansion of π with respect to calibration parameter updates $\delta g_1, \ldots, \delta g_m$

$$\begin{aligned} &(\mathcal{I} \circ \pi)(g_1 + \delta g_1, \ldots, \delta g_m + \delta g_m) \approx \\ &(\mathcal{I} \circ \pi)(g_1, \ldots, g_m) + D\mathcal{I} \left(\sum_{i=1}^{m} \frac{\mathrm{d}\pi}{\mathrm{d}g_i} \delta g_i \right), \end{aligned} \qquad (3)$$

where $D\mathcal{I}$ denotes the Jacobian of $\mathcal{I}$ with respect to the projection parameters. Plugging this into (2) yields the following functional

$$\int_{\beta(S)} \left\| \mathcal{I} \circ \pi + D\mathcal{I} \left(\sum_{i=1}^{m} \frac{\mathrm{d}\pi}{\mathrm{d}g_i} \delta g_i \right) - T \right\| \det(D\pi) \, \mathrm{ds}. \qquad (4)$$

Now, (4) is minimized with respect to the calibration updates $\delta g_i, i = 1, \ldots, m$. To compute the derivatives of π, we make use of the exponential parametrization. For technical details, we refer to [21].

2.4. Limitations of the Direct Approach

The above approach is a generalization of the model in [21]. While also relying on the reprojection error, the main differences are the use of a L_1-norm and a super-resolution texture estimation. However, experiments indicate that local minimization of this highly non-convex optimization problem gives rise to suboptimal solutions and often does not lead to substantial improvements in the estimated camera parameters. In the next section, we will provide reasons for this shortcoming and propose a decoupling strategy which leads to a considerably more robust calibration method.

3. Decoupling Photometry and Geometry

A closer look at functional (1) reveals that the camera parameters are being estimated so as to minimize the *photometric* error between modeled and observed texture. Interestingly, this is in sharp contrast to the established bundle adjustment approach for accurate camera calibration which aims at minimizing the *geometric* error between observed points and the corresponding back-projected 3D points. In particular, the problems of estimating point correspondences and minimizing the geometric error are treated separately. This is important because, upon gradual improvement of the camera parameters, the geometric error is likely to decrease, whereas – in particular for high-resolution textures – the photometric error is more likely to oscillate (rather than decrease), thus leading to bad convergence of algorithms based on pure photometric criteria. In other words, incorporating a geometric measure reduces the number of local minima and gives clearer evolution directions. Furthermore, the success of established tools like bundle adjustment indicates that for accurate camera calibration based on high-resolution textures, one should separate the algorithmic problems of correspondence estimation and calibration.

In the following, we will demonstrate that a relaxation scheme for minimizing energy (1) provides exactly the desired solution to the above problem.

3.1. Decoupled Energy

To achieve the goal of decoupling the two subproblems, we introduce for each camera an additional displacement field $v : \Omega \to \mathbb{R}^2$ defined on the image plane, which resembles the error in the projection. The key observation is that the image formation model shall be matched exactly if each point is displaced by v to account for the error. At the minimum, v should of course be as small as possible. By reformulating a single term of energy (1) in terms of this new displacement field, we arrive at

$$E(\pi, v) = \int_S \|\mathcal{I}(x + v(x)) - \mathcal{T}(x)\| \, \mathrm{dx} + \alpha \|v\|_{1,1} , \quad (5)$$

where $\alpha > 0$ is a weighting parameter and $\mathcal{T} := b * (T \circ \beta)$ is the appearance of the object using the current calibration and texture. The model in (5) can be interpreted as a relaxation of energy (1), since if $\alpha \to \infty$, the displacement v is forced to be zero and we arrive at the original solution for π. To regularize v, we choose the Sobolev norm $\|\cdot\|_{1,1}$, since in order to stabilize the solution, it is necessary to not only penalize large displacements but also large changes of the displacements, i.e. the first derivative.

Comparing the energy functional in (5) to the methodology of sparse calibration methods, we observe that the proposed dense formulation allows to propagate neighboring information by regularizing the underlying displacement field v. Thereby, well-textured regions prevail, while homogeneous regions give rise to displacements close to zero. It should be noted that the formulation in (5) can easily be "sparsified" by integrating in the first term a weighting function $w : \Omega \to \{0, 1\}$ (or a relaxed version $w : \Omega \to [0, 1]$) which accounts for the reliability of the respective pixel measurement. Yet, we found out that this is usually not necessary in practice.

One can observe that the energy model in (5) is expressed in terms of two conceptually different arguments – the projection π and the image-based displacement field v. Thus, a straightforward way to accomplish the minimization is to split the functional into two parts $E_1(v) + E_2(\pi)$ and to optimize iteratively E_1 with respect to v and E_2 with respect to π.

3.2. Computing the Geometric Reprojection Error

In the proposed formulation, the first part of the energy is comprised of the data term and the derivative part of the Sobolev norm

$$E_1(v) = \int_S \|\mathcal{I}(x + v(x)) - \mathcal{T}(x)\| + \alpha \|Dv(x)\| \, \mathrm{dx}. \quad (6)$$

It resembles a TV-L_1 optical flow model and can be minimized with the algorithm detailed in [24].

This step allows to compute a geometric error v, while minimizing the photometric error.

In order to make the approach applicable to the non-Lambertian case, i.e. to make it robust to illumination changes, in a preprocessing step we perform classical normalization of the image intensities. In particular, we transform the images $\mathcal{I}$ and $\mathcal{T}$ to achieve a contrast invariant form of the error term. Denoting by $\mu(\mathcal{I}, x)$ the local mean of the image $\mathcal{I}$ at x and by $\sigma(\mathcal{I}, x)$ the respective standard

deviation, the transformation for $\mathcal{I}$ is defined as

$$\frac{\mathcal{I}(x) - \mu(\mathcal{I}, x)}{\sigma(\mathcal{I}, x)} \tag{7}$$

and analogously for $\mathcal{T}$. In our experiments, we applied this procedure on 5×5 patches.

3.3. Optimization w.r.t. the Projection

The idea behind the second part of the energy is to explain the error v in the projection by the error in the calibration - that is, we want to adapt π such that it leads to a reduction of the observed error v in the next iteration. Comparing the total energy (5) and the first part (6), we observe that the missing term for the second part is the L_1-norm of v. We now make the dependence on the projection explicit.

The intuition behind the displacement field $v(x)$ obtained at a point x in the silhouette is that the point should instead be projected onto $x + v(x)$ for an optimally photo-consistent match, see figure 2. We can achieve this goal by rewriting the second part of the energy in terms of π and then updating the projection accordingly. Therefore, we introduce a mapping $p : \beta(S) \to \Omega$, which assigns the (possibly erroneous) projection to each point on the visible part of the surface. The defining equation is $(p \circ \beta)(x) = v(x) + x$. Thus, p could be regarded as a generalized projection taking the correction field v into account. Note that it is kept fixed during the subsequent optimization of π. In order to match the desired updated projection to the observed error, note that for a visible point s on the surface, we have $v \circ \pi(s) = p(s) - \pi(s)$, since $\beta \circ \pi(s) = s$. Using this identity, we can transform the second part of the energy up to the surface and rewrite it as

$$E_2(\pi) = \int_{\Omega} \|v(x)\| \, \mathrm{dx} = \int_{\beta(S)} \|p - \pi\| \det(D\pi) \, \mathrm{ds}. \tag{8}$$

This equation resembles the standard discrete bundle adjustment process in which $\sum_j (x_j - \pi(s_j))^2$ is being minimized, where s_j are the different discrete 3D points and x_j – the corresponding projections. Note that we have an additional term which accounts for the foreshortening of the surface under perspective projection. In fact, modeling this important aspect in a rigorous way is only possible in a dense formulation.

We minimize E_2 in the same way as already described for the energy in paragraph 2.3. Plugging in the Taylor expansion of the projection around the currect values of the projection parameters, we again end up with an energy which we can minimize with respect to the calibration parameter updates $(\delta g_1, \ldots, \delta g_m)$. The energy (8) written in terms of these updates appears in step 5 of the complete algorithm, which is summarized in table 3.

INPUT: 3D model, texture model T, input images I_k, camera parameters g_i^k

1: **for all** k **do**
2: **repeat**
3: compute $T \circ \beta_k$
4: compute optical flow v between $T \circ \beta_k$ and I_k by minimizing $E_1(v)$
5: compute $\left\{\delta g_i^k\right\}_i$ to minimize $\int_{\beta_k(S)} \left| \sum_{i=1}^{m} \frac{\mathrm{d}\pi}{\mathrm{d}g_i} \delta g_i - v \circ \pi_k \right| \det(D\pi) \, \mathrm{ds}$
6: **for all** i **do**
7: $g_i^k \leftarrow g_i^k + \delta g_i^k$
8: **end for**
9: **until** convergence
10: **end for**
11: **RETURN** $\left\{\delta g^k\right\}_k$

Figure 3. *Algorithm for decoupled variational calibration.*

4. Experiments and Results

We evaluate the performance of the proposed calibration approach by examining the improvements of the estimated texture map as well as the reconstructed 3D geometry.

4.1. Direct vs. Decoupled Approach

In figure 4, we compare the results of the decoupled method and the direct photometric solution described in section 2.3. We used an image sequence capturing a bunny figurine consisting of 33 views with manually distorted calibration parameters[1]. It can be observed that while the photometric error converges to a similar value for both methods, the geometric error, which is represented by the optical flow, converges to a substantially smaller value with our algorithm than with the direct approach. Furthermore, the decoupled model leads to less outliers, a faster convergence rate and more resilience to local minima. Contrary to sparse feature-point based methods, the proposed formulation also has the advantage of not favoring any particular regions depending on the number of salient points. This is important to avoid scene-specific accumulation effects.

4.2. Improving the Texture

A first experiment, which shows that our framework indeed improves the camera parameters, is to compute a texture map for a fixed geometric model before and after applying the proposed calibration optimization (see figure 5). In order to be independent from the particular approach for texture estimation, we show comparisons for the naive averaging method as well as the super-resolution method of [6].

[1]The image sequences in figure 4 and 6 are publicly available on our webpage, `http://cvpr.in.tum.de/research/datasets`

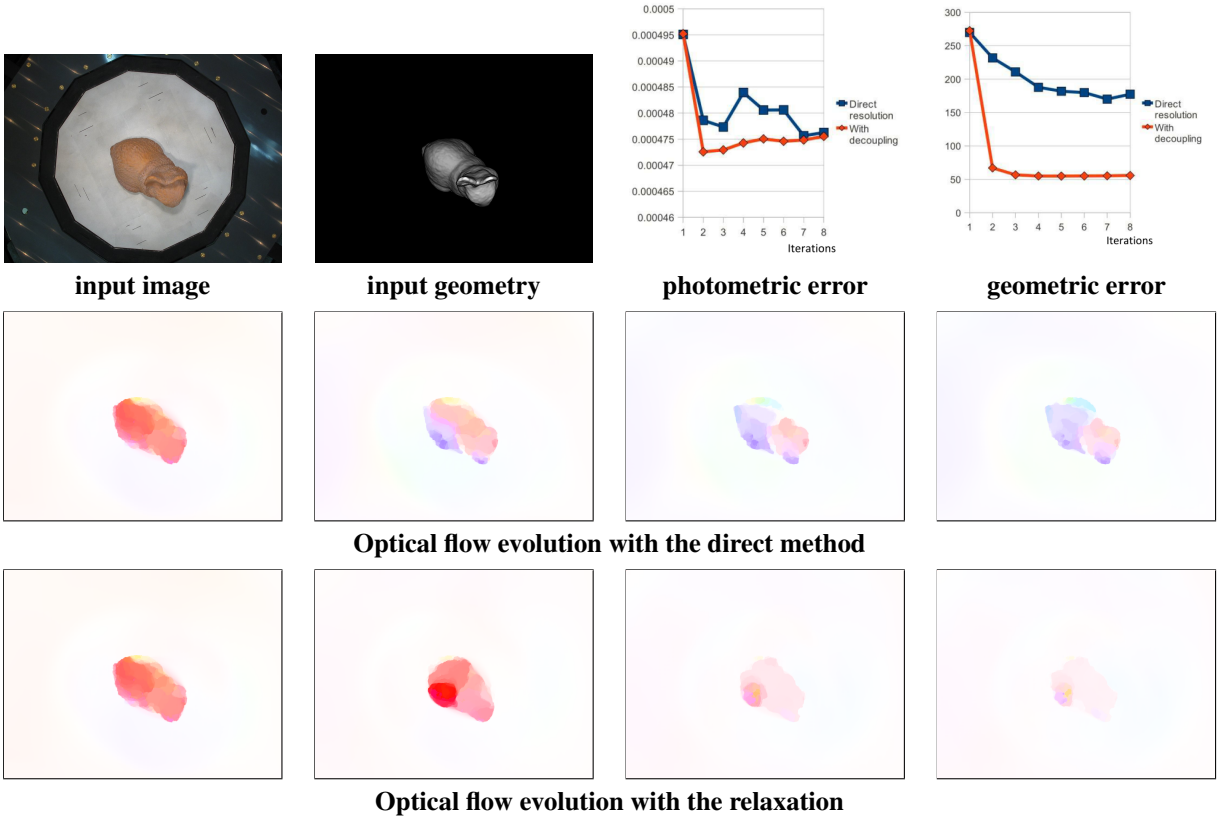


Figure 4. Comparison of the direct method and the method improved by our relaxation. While the photometric errors converge to similar values in both cases, the geometric error is much lower with the proposed decoupling method. This can be understood by looking at the optical flow evolution, where we can see that the direct method gets stuck in a local minimum. In our proposed scheme the magnitude of the flow, which represents the geometric reprojection error, is overall much lower and converges steadily.

It can be observed that the simple color averaging technique produces quite blurry results. Nevertheless, the proposed calibration procedure leads to a visible enhancement of the texture pattern.

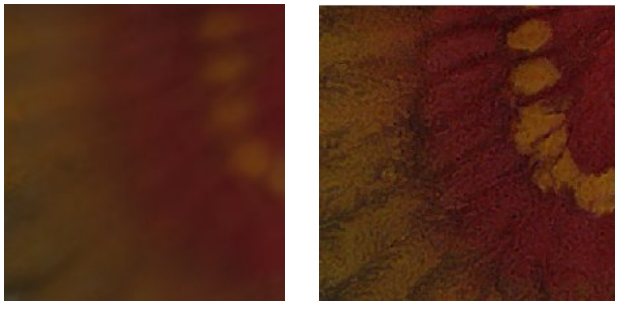
The improvements are even more notable when applying the super-resolution framework. In this case, visual artifacts are removed to obtain a high-quality texture map. Note that even some of the fine-scale texture details are clearly visible. Moreover, with the improved calibration of the images approximately 10 times less iterations of the super-resolution algorithm were necessary to converge to the visualized result. This is not surprising, since a better calibration gives rise to more accurate costs and thus – more precise derivative directions which guide the optimization process.

4.3. Improving the Geometry

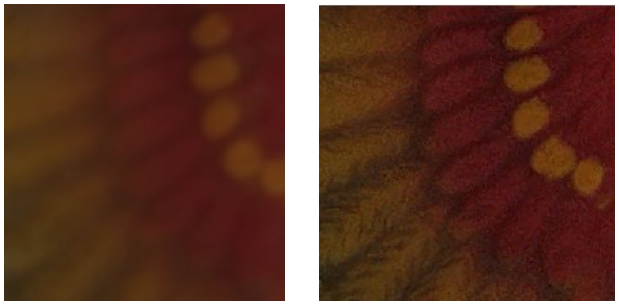
By applying the proposed variational calibration approach, we observed substantial improvements also in the reconstructed 3D models. Generally, the changes are more notable for small image sequences, since for large datasets calibration inaccuracies sum up and balance each other. Note, however, that a reduced number of images pose a great challenge for classical sparse calibration methods due to the large baselines which considerably exacerbate the matching process.

Our first test sequence captures a bird figurine from 21 vantage points. The original calibration was obtained by applying a classical LED-based calibration procedure. A traditional visual hull computation with the original and refined camera parameters already shows some small but considerable improvements, see figure 6. This is clearly visible at the leg, where miscalibrations cause notable artifacts.

In our second test case, we explore the sensibility of a multi-view stereo algorithm [10] on calibration refinement, see figure 7. The dataset consists of 16 image of a temple replicate and is used in the well-known Middlebury multi-view stereo evaluation challenge as “templeSparseRing” (see [14]). Clearly, the reconstruction with the original camera parameters exhibits more irregularity than the reconstruction with the parameters optimized with our algorithm. This is additionally confirmed by a quantitative



original calibration



improved calibration

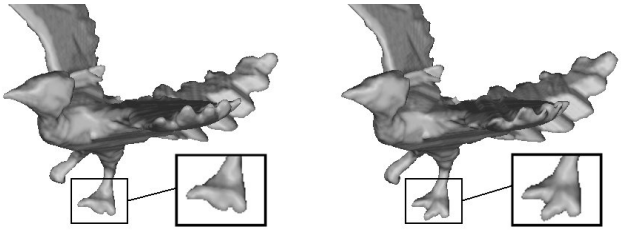
Figure 5. Texture improvements after refining the camera parameters with the proposed approach for the bird data set (see fig. 6). The original calibration was obtained by applying a classical LED-based calibration procedure. Visualized are zoomings of an average texture map (left) and a super-resolution texture map (right). Note the enhancement of the texture pattern as well as the removal of visual artifacts obtained with the refined calibration parameters.

evaluation. The percentage of points within a band of 2 millimeters was increased by 2%, and the mean accuracy of the 99% best points was improved by 1.45 millimeters, whereas the percentage of points within a band of 1.25 millimeters was increased by 2.7%, and the mean accuracy of the 90% best points was improved by 0.16 millimeters. These results point out the strong reduction of the number of outliers, as well as a general improvement in accuracy. This conclusion is in agreement with the observations in [4].

5. Conclusion and perspective

We have proposed a novel method to perform variational camera calibration in a spatially dense setting. Using a relaxation technique, which allows to decouple the estimation of point correspondences and camera parameters, we break down the original minimization problem into two distinct subproblems and solve them alternately. The first subproblem is a dense correspondence estimation which is addressed by an optical flow algorithm, the second – a camera parameter estimation which resembles a continuous dense form of bundle adjustment. Experiments demonstrate that the decoupling strategy leads to a more accurate result than a direct minimization approach. The refined camera parameters provided by our method lead to a significant improvement of the estimated super-resolved texture maps as well as the reconstructed 3D models.

input images (3/21)



original calibration **improved calibration**

Figure 6. On the bird dataset, refining the calibration parameters with the proposed approach leads to small but substantial improvements in the subsequently computed visual hull. The original calibration was obtained by applying a classical LED-based calibration procedure.

As a future work, it would be interesting to explore if the proposed formulation could be extended by incorporating further established concepts from sparse calibration methods like epipolar constraints to additionally increase the robustness of the approach. Moreover, a generalization unifying camera calibration, texture and geometry estimation in a single framework seems to be challenging but quite appealing.

References

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, 2002. 2

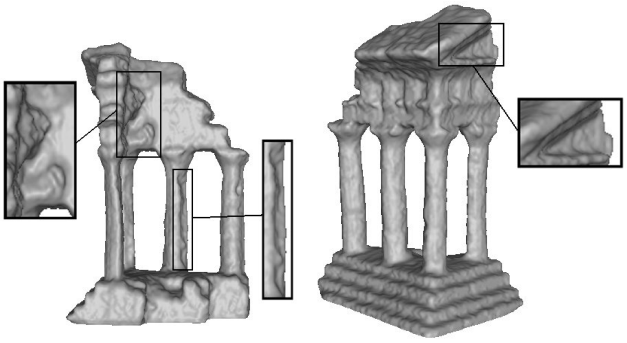
[2] C. H. Esteban, F. Schmitt, and R. Cipolla. Silhouette coherence for camera calibration under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):343–349, 2007. 2

[3] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In *Readings in computer vision: issues, problems, principles, and paradigms*, pages 726–740. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987. 2

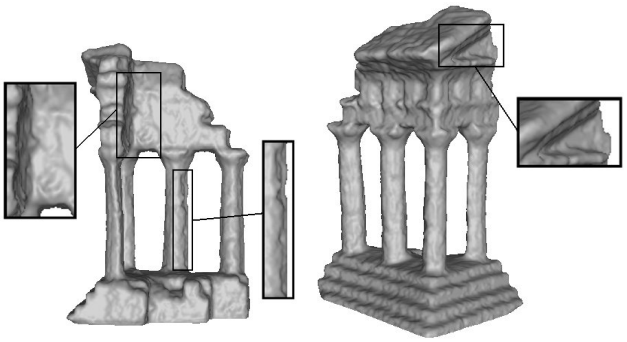
[4] Y. Furukawa and J. Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. *International Journal of Computer Vision*, 84:257–268, 2009. 2, 7

[5] B. Goldluecke and D. Cremers. A superresolution framework for high-accuracy multiview reconstruction. In *Pattern Recognition (Proc. DAGM)*, 2009. 1

input images (4/16)



original calibration



improved calibration

Figure 7. Multi-view stereo can be improved by using the proposed calibration optimization. The reconstruction of [10] on "templeSparseRing" with the parameters fournished for the Middlebury challenge exhibits more irregularity than the reconstruction with the parameters optimized with our algorithm. Quantitative results are also clearly improved.

[6] B. Goldluecke and D. Cremers. Superresolution texture maps for multiview reconstruction. In *IEEE International Conference on Computer Vision (ICCV)*, 2009. 2, 3, 5

[7] C. G. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fouth Alvey Vision Conference*, pages 147–152, 1988. 2

[8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004. 1

[9] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 225 –234, 2007. `http://ewokrampage.wordpress.com/`. 1

[10] K. Kolev, M. Klodt, T. Brox, and D. Cremers. Continuous global optimization in multiview 3D reconstruction. *International Journal of Computer Vision*, 84(1):80–96, 2009. 3, 6, 8

[11] K. Kolev, T. Pock, and D. Cremers. Anisotropic minimal surfaces integrating photoconsistency and normal information for multiview stereo. In *European Conference on Computer Vision (ECCV)*, Heraklion, Greece, September 2010. 3

[12] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 2

[13] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:756–777, 2004. 2

[14] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 519–528, 2006. 1, 6

[15] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: Exploring image collections in 3D. In *Proceedings of the ACM SIGGRAPH*, 2006. `http://phototour.cs.washington.edu/bundler/`. 1

[16] S. Soatto, A. J. Yezzi, and H. Jin. Tales of shape and radiance in multiview stereo. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 974–981, October 2003. 3

[17] F. Sroubek, G. Cristobal, and J. Flusser. A unified approach to superresolution and multichannel blind deconvolution. *IEEE Transactions on Image Processing*, 16(9):2322–2332, 2007. 2

[18] C. Strecha, W. von Hansen, L. V. Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2008. 1

[19] E. Tola, V. Lepetit, and P. Fua. A fast local descriptor for dense matching. In *Conference on Computer Vision and Pattern Recognition*, Alaska, USA, 2008. 2

[20] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment – a modern synthesis. In *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer, 2000. 2

[21] G. Unal, A. Yezzi, S. Soatto, and G. Slabaugh. A variational approach to problems in calibration of multiple cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1322–1338, 2007. 2, 3, 4

[22] K.-Y. K. Wong and R. Cipolla. Reconstruction of sculpture from its profiles with unknown camera positions. *IEEE Transactions on Image Processing*, 13:381–389, 2004. 2

[23] A. Yezzi and S. Soatto. Structure from motion for scenes without features. In *Proc. International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 171–178, 2003. 2

[24] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L1 optical flow. In *Proceedings of the 29th DAGM conference on Pattern recognition*, pages 214–223, 2007. 4