Can a Fully Unconstrained Imaging Model be Applied Effectively to Central Cameras?

Filippo Bergamasco, Andrea Albarelli, Emanuele Rodolá, Andrea Torsello

Dipartimento di Scienze Ambientali, Informatica e Statistica

Università Ca' Foscari Venezia - via Torino, 155 - 30172 Venice Italy

`http://www.dais.unive.it`

Abstract

Traditional camera models are often the result of a compromise between the ability to account for non-linearities in the image formation model and the need for a feasible number of degrees of freedom in the estimation process. These considerations led to the definition of several ad hoc models that best adapt to different imaging devices, ranging from pinhole cameras with no radial distortion to the more complex catadioptric or polydioptric optics. In this paper we propose the use of an unconstrained model even in standard central camera settings dominated by the pinhole model, and introduce a novel calibration approach that can deal effectively with the huge number of free parameters associated with it, resulting in a higher precision calibration than what is possible with the standard pinhole model with correction for radial distortion. This effectively extends the use of general models to settings that traditionally have been ruled by parametric approaches out of practical considerations. The benefit of such an unconstrained model to quasi-pinhole central cameras is supported by an extensive experimental validation.

1. Introduction

A camera imaging model describes mathematically the optical process that drives photons to the sensitive surface inside the capturing device. In practice, this boils down to the mapping between the pixels of the CCD or CMOS of the camera and a set of corresponding rays in 3D space. The goal of any calibration method is to build such map by finding the optimal parameters for the model. In turn this results in tight coupling of models and calibration procedures, since any calibration approach is designed for a specific camera type.

The simplest formalization of the imaging process is the pinhole camera. This basic model can be calibrated by solving a linear system that relates the coordinates of refer-ence points in the scene with their projections on the image plane [5]. Unfortunately, no real camera behaves exactly like an ideal pinhole. In fact, in most cases, at least the distortion effects introduced by the lens should be accounted for [19]. Any pinhole-based model, regardless of its level of sophistication, is geometrically unable to properly describe cameras exhibiting a frustum angle that is near or above 180 degrees. For wide-angle cameras, several different parametric models have been proposed. Some of them try to modify the captured image in order to follow the original pinhole behavior [8, 4]. Others go trough a totally different path by introducing novel image formation models [9]. Also catadioptric systems have been widely covered in the literature, with a large selection of models and calibration methods [1, 13].

Given this proliferation of different models, the desire for an unifying approach is quite natural. The most general imaging model, that associates an independent 3D ray to each pixel, would in principle be able to describe any kind of imaging system, regardless of the optical path that drives each ray to the sensitive elements of the device. However, the complete independence of millions of rays makes its calibration a daunting task as each of them needs several 2D to 3D correspondences to be properly constrained.

This problem was first addressed in [6], where unconstrained rays (here called raxels) are calibrated exploiting their intersections with a target (an encoded laptop monitor) that moves along a translating stage. Such intersections are identified by means of Gray coding and are evaluated as the average over several different shots. This approach is somewhat limited by the fact that the pose of the calibration planes must be known (i.e. the method depends on the accuracy of the translating stage), and the paper does not assess the accuracy of the obtained calibration.

A method for the calibration of the general model and unknown poses is proposed in [18]. The authors discuss both the case of non-central and perspective camera, however, in the latter case, the parametrization process has proven to be rather complicated when using planar calibra-

tion objects [17]. Further, the paper does not describe a specific systematic setup for gathering 2D to 3D correspondences for all the camera rays and the experimental evaluation is performed qualitatively in a subset of the imaging sensor.

The practical usage of multiple grids for calibrating generic non-perspective device is investigated in [14]. Again, the method is designed and well-suited for catadioptric, spherical, multiview and other types of non-central cameras.

In general, the literature has deemed the unconstrained model and related calibration procedures a last resort to be adopted only when traditional approaches fail due to either geometrical or methodological issues. For this reason the pinhole model, augmented with a proper distortion correction, dominates the application landscape whenever its use is feasible. With this paper we explore the opposite direction. Specifically, we ask ourselves (and the reader) whether even the most traditional perspective camera could benefit of the adoption of a non-parametric model. For this to be the case, the calibration must be both effective and reasonably easy to perform. In the following sections we briefly describe our generic model (which is indeed pretty standard) and we introduce a practical calibration method. The impact on the calibration accuracy with respect to the pinhole model is evaluated with a wide set of experiments. Finally, aspects and implications related to the use of an unconstrained model with common computer vision tasks are discussed.

2. Imaging Model and Calibration

In this paper we propose to use a fully unconstrained camera model even for central cameras. In the proposed model each pixel is associated with the light ray direction from the object to where the ray hits the optics, rendering completely irrelevant how the optics bend the light to hit the CCD. This ray can be formalized as a line in the Euclidean space which, in the unconstrained model, is independent on the lines assigned to the other pixels and completely free with respect to direction and position. Under these assumptions, pixels cease to hold a precise geometrical meaning and they become just indexes to the imaging rays, having completely hidden the path that the ray has to go through inside the optics to hit the right cell in the CCD.

In what follows, index i ranges over camera pixels. The ray associated with camera pixel i can be written as $\mathbf{r}_i = (\mathbf{d}_i, \mathbf{p}_i)$, where $\mathbf{d}_i, \mathbf{p}_i \in \mathbb{R}^3$ represent direction and position of the ray respectively (see Figure 1). These vectors satisfy $||\mathbf{d}_i|| = 1$, (normalized direction) and $\mathbf{d}_i^T \mathbf{p}_i = 0$ (orthogonal position vector). Any point $\mathbf{x}$ in the ray $\mathbf{r}_i$ satisfies the parametric equation $\mathbf{x} = \mathbf{d}_i t + \mathbf{p}_i$ for some $t \in R$.

This model has 4 degrees of freedom per pixel, resulting in several million parameters to be estimated for current cameras, a dimensionality that is beyond the possibilities of the most commonly used calibration processes. Note, however, that the ray independence assumptions allows for (conditionally) independent estimation of each ray, allowing us to measure the convergence of the estimate as a measure of per-pixel observation. One problem with the commonly used target-based calibration systems (be them chessboard-based, dot-based, or based on any other pattern) is that they provide sparse localization points, resulting in rather low numbers of per-pixel observations. We argue that it is this situation that forces the use of a low dimensional imaging model as the unconstrained model would not converge for every ray, even with a very large number of poses.

We propose to solve this problem by providing dense localization on the target, thus resulting in one observation per pixel in each pose of the calibration target. This dense calibration target is obtained through the use of structured light pattens on a normal LCD display, allowing us to assign to each camera pixel the 2D coordinate in the target planar reference frame of the location where the ray associated to the pixel hits the target. In particular, we use phase coding with the number-theoretical phase unwrapping approach presented by Lilienblum and Michaelis [12] to encode both the horizontal and vertical coordinate of each pixel of the target display. Each camera pixel integrates this signal over all the incoming rays incident that CCD cell, resulting in a high precision sub-pixel localization of the average incident ray and thus a localization in the target's surface with a precision that is sub-pixel with respect to the target's pixel dimension.

Let index s range over the calibration shots, with $\boldsymbol{\Theta}_s = (\mathbf{R}_s, \mathbf{t}_s)$ we denote the pose parameters of the calibration target in shot s, transforming the target's coordinate system onto the camera coordinates. The $\mathbf{u}_s$, $\mathbf{v}_s$, and $\mathbf{n}_s$ base vectors of the target coordinate system expressed in the camera coordinate system correspond to the first, second and third columns of $\mathbf{R}_s$ respectively, i.e., $\mathbf{R}_s = (\mathbf{u}_s \mathbf{v}_s \mathbf{n}_s)$.

Further, let $\mathbf{Co}_i^s \in \mathbb{R}^2$ denote the code (target 2D location) measured at camera pixel i in shot s, while with $\mathbf{Ce}(\mathbf{r}_i | \boldsymbol{\Theta}_s) \in \mathbb{R}^2$ we denote the expected code at pixel i, given ray $\mathbf{r}_i$ and target pose $\boldsymbol{\Theta}_s$ (see Figure 1). Ignoring possible refraction effects on the monitor's surface this corresponds simply to the surface coordinates (u, v) of the intersection between the ray and the target plane. To add the effects of refraction, we need to add a correction term accounting for Snell's law. For the moment we will concentrate on the refraction-less case, and add the refraction term and analyze its effects on Subsection 2.4. In the refraction-less case, we have:

$$
\mathbf{Ce}(\mathbf{r}_i | \boldsymbol{\Theta}_s) = \mathbf{P_{uv}} (\mathbf{d}_i t_{\text{int}} + \mathbf{p}_i) =
$$
$$
(\mathbf{u}_s \mathbf{v}_s)^T \left(\frac{\mathbf{n}_s^T (\mathbf{t}_s - \mathbf{p}_i)}{\mathbf{n}_s^T \mathbf{d}_i} \mathbf{d}_i + (\mathbf{p}_i - \mathbf{t}_s) \right) , \quad (1)
$$

Figure 1. Schema of the general camera model and calibration target described in this paper. Note that the Mahalanobis distance between observed and expected code is equal to the 3D distance of observed code to the ray.

where $\mathbf{P_{uv}}$ denotes the projection onto the (u, v) planar coordinates, and

$$t_{\text{int}} = \frac{\mathbf{n}_s^T(\mathbf{t}_s - \mathbf{p}_i)}{\mathbf{n}_s^T \mathbf{d}_i} \qquad (2)$$

is the intersecting parameter for the equation of ray $\mathbf{r}_i$, i.e., the value such that $\mathbf{d}_i t_{\text{int}} + \mathbf{p}_i$ lies on the target plane.

Under this setup the calibration process can be cast as a joint estimation of the $\mathbf{r}$ and $\boldsymbol{\Theta}$ parameters.

2.1. Least Squares Formulation

We express the calibration process as a generalized least squares problem [10]. Generalized least squares is a technique for estimating the unknown parameters from a (non-)linear model in presence of heteroscedasticity, i.e., where the variance of the observations are unequal and/or correlated. In such a situation ordinary least squares can be statistically inefficient, while the general least squares estimator provides an efficient estimator. The approach accounts for heteroscedasticity by normalizing the residuals through the inverse of the covariance of the measurements $\boldsymbol{\Sigma}$, thus minimizing the sum of the squared Mahalanobis lengths of the residuals:

$$\sum_k (y_k - f(x_k, \theta)) \boldsymbol{\Sigma}_k^{-1} (y_k - f(x_k, \theta))^T . \qquad (3)$$

Let $\varepsilon_i^s = \mathbf{Co}_i^s - \mathbf{Ce}(\mathbf{r}_i | \boldsymbol{\Theta}_s)$ be the code residuals, then the generalized least squared estimate of the rays and pose parameters $\mathbf{r}, \boldsymbol{\Theta}$ is:

$$(\hat{\mathbf{r}}, \hat{\boldsymbol{\Theta}}) = \underset{\mathbf{r}, \boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{i,j,s} (\varepsilon_i^s)^T (\boldsymbol{\Sigma}_i^s)^{-1} \varepsilon_i^s , \qquad (4)$$

where $\boldsymbol{\Sigma}_i^s$ is the (conditional) error covariance matrix under the given pixel-pose combination.

In this context the main source of heteroscedasticity derives from the directional correlation of code errors when rays hit the target plane at an angle. In fact, let ϕ be the angle between the ray direction $\mathbf{d}$ and the normal to the target $\mathbf{n}$, and let $\mathbf{r}_i^{\parallel s} = \frac{(\mathbf{u}_s \mathbf{v}_s)^T \mathbf{d}_i}{||(\mathbf{u}_s \mathbf{v}_s)^T \mathbf{d}_i||}$ be the direction of the ray projected onto the target's planar coordinates, then measurement errors will be amplified by a factor of $\frac{1}{\cos^2 \phi}$ along $\mathbf{r}_i^{\parallel s}$ due to the effects of the grazing angle, while will remain unchanged along the orthogonal direction. Hence, we obtain

$$\begin{aligned} (\boldsymbol{\Sigma}_i^s)^{-1} &= I + (\cos^2 \phi_i^s - 1) \mathbf{r}_i^{\parallel s} (\mathbf{r}_i^{\parallel s})^T \\ &= I - (\mathbf{u}_s \mathbf{v}_s)^T \mathbf{d}_i \mathbf{d}_i^T (\mathbf{u}_s \mathbf{v}_s) . \qquad (5) \end{aligned}$$

In the standard pinhole model the effect of eliminating the correlation of the errors is obtained by re-projecting the residuals onto the image plane, i.e., by minimizing the reprojection error rather than the planar displacement over the target. In this sense, normalizing over the inverse error covariance is as close as one can get to the minimization of the reprojection (or geometrical) error in an unconstrained system that loses any direct geometrical connection between rays and pixels. It is worth mentioning that the reprojection error in the pinhole model accounts for another source of heteroscedasticity, i.e., a change of scale in the covariance as the targets moves away from the camera as a consequence of foreshortening. In the unconstrained model this would imply an unknown scale term on the covariance $\boldsymbol{\Sigma}$ that would require inter-ray interaction in order to be estimated, complicating the least square formulation and hindering our ability to estimate the parameters effectively and/or with reasonable computational effort. For this reason in the formulation we are ignoring the depth-related change of scale in the error variance. We note, however, that this effect is relatively limited, since it would induce a variation in the scale of the error covariance proportional to $\frac{\Delta z}{\bar{z}}$ (depth variability over average depth) which in our setup reduces to a variability within approximately $\pm 10\%$ of the average.

To optimize the least squares formulation efficiently, we make use of the conditional independence of the ray parameters $\mathbf{r}$ given the poses $\boldsymbol{\Theta}$ and of the poses given the rays. We do this by performing a two-step optimization process in which we alternatively optimize all the ray parameters in parallel keeping the pose parameters fixed, and then optimize the poses keeping the rays fixed. This way, the large scale estimation part, i.e., the optimization of the ray parameters, becomes massively parallel and can be computed very efficiently in GPU. In our experiments the optimization process is initialized with the normal pinhole model with polynomial radial distortion.

2.2. Ray Calibration

As we fix the pose parameters, all the rays depend only on the observed coordinates $\mathbf{Co}$ associated with each of them and can be estimated independently. Further, with the pose parameters at hand, these observed 2D coordinates can be transformed into 3D points in the camera coordinate frame. As can be seen in Figure 1, given a ray $\mathbf{r}$ intersecting the target plane at 2D coordinate $\mathbf{Ce}$, we can divide the residual $\varepsilon = \mathbf{Ce} - \mathbf{Co}$ into the orthogonal vectors $\varepsilon^\parallel = \mathbf{Ce} - \mathbf{Co}^\parallel$ and $\varepsilon^\perp = \mathbf{Co}^\parallel - \mathbf{Co}$, where $\varepsilon^\parallel$ is parallel to $\mathbf{r}^\parallel$. Clearly, since $\varepsilon^\perp$ is orthogonal to the plane spanned by $\mathbf{d}$ and $\mathbf{n}$, the point in $\mathbf{r}$ closest to $\mathbf{Co}$ is also the one closest to $\mathbf{Co}^\parallel$. Further, let $\mathbf{h}$ be this point, we have

$$||\mathbf{h} - \mathbf{Co}||^2 = ||\mathbf{h} - \mathbf{Co}^\parallel||^2 + ||\varepsilon^\perp||^2 \,. \qquad (6)$$

It is easy to show that, $||\mathbf{h} - \mathbf{Co}^\parallel|| = \cos\phi ||\varepsilon^\parallel||$, where ϕ is the angle between $\mathbf{d}$ and $\mathbf{n}$. Hence, the squared distance between $\mathbf{r}$ and $\mathbf{Co}$ equals

$$d^2(\mathbf{r}, \mathbf{Co}) = \cos^2\phi ||\varepsilon^\parallel||^2 + ||\varepsilon^\perp||^2 = \varepsilon^T \boldsymbol{\Sigma}^{-1} \varepsilon \,, \qquad (7)$$

thus the generalized least squares formulation with respect to the target coordinates corresponds to the standard linear least squares with respect to the 3D points associated with each ray. The linear least squares problem is then solved by a ray with parametric equation $\bar{\mathbf{x}} + \mathbf{w}t$, where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ is the barycenter of the observed 3D points, and $\mathbf{w}$ is the eigenvector of their covariance matrix corresponding to the smallest eigenvalue.

2.3. Estimation of the Poses

The second step in the alternating calibration process is the optimization of the target poses keeping the rays fixed. Also in this step we can make use of a conditional independence; in fact, with the rays fixed, the pose parameters from different shots become independent and can be optimized in parallel. Further, just like in the ray calibration step, we make use of the equivalence between the Mahalanobis distance over the 2D core errors $\mathbf{Ce} - \mathbf{Co}$ with the 3D Euclidean distance between observed position and ray. In this situation the estimation of pose $\boldsymbol{\Theta}_s$ reduces to the search for the rigid transformation that minimizes the distance between the (transformed) observations $\boldsymbol{\Theta}_s \mathbf{Co}_i^s$ and the ray $\mathbf{r}_i$. This minimization problem shares several similarities with surface alignment, where we seek to find the rigid transformation that minimizes the Euclidean distance between a set of points on a surface to points on another surface. It is not surprising, then, that a straightforward modification of the Iterative Closest Point (ICP) algorithm [2] can be used to find a (local) optimum for our problem. The modification lies in how the closest points are sought: In our context instead of searching for the closest point in a

single given 2D surface common for all the points, we select the closest point in the unique 1D ray associated with the given observation. The iterative selection of all the closest points for all the pixels in each shot and application of Horn's alignment algorithm [7] converges to a locally optimal pose.

Clearly, given the generality of the imaging model, the optimization problem is not convex, thus we cannot guarantee a global optimum like in the case of the pinhole model. However, under the assumption of a central quasi-pinhole camera, we can obtain a very good initialization and be confident about the quality of the final solution found.

2.4. Accounting for Refraction

In our formulation we have ignored the possible effects of refraction on the monitor's surface.

In [15] it was shown that refraction had a small but noticeable effect when calibrating using LCD displays, so we extended the least square formulation to incorporate Snell's law of optical refraction in order to assess and correct its effects. To this end, we added two global parameters λ and μ representing respectively the (inverse) refractive index between air and the transparent layer in front of the LCD, and the depth of the layer. According to Snell's law, a ray $\mathbf{r}$ hitting the surface at an angle ϕ with the normal $\mathbf{n}$, will be refracted inside the transparent layer at an angle ψ satisfying $\sin\psi = \lambda \sin\phi$, hitting the reflective layer at target coordinates $\mathbf{Ce} + \Delta\mathbf{Ce}$ with

$$\Delta\mathbf{Ce} = \mathbf{r}^\parallel \mu \tan\psi = \mu \frac{\lambda \sin\phi}{\sqrt{1 - \lambda^2 \sin\phi}} \,. \qquad (8)$$

The addition of refraction adds a non-linearity in both the ray and pose estimation that breaks the conditional independence assumption at the basis of our approach. We solve this by adopting a fixed point approach, reiterating the least squares estimations (both in the ray and pose estimation phases) using the refraction shift $\Delta\mathbf{Ce}$ computed based on the previous rays and poses.

Figure 2 shows the effect of the refraction parameters on the final root mean squared error (RMS) of the calibrated camera. From the plot we can see clearly that the minimum is attained in the refraction-less case ($\mu = 0$ or $\lambda = 0$), thus pointing to a negligible effect of refraction for the unconstrained model as opposed to what was reported in [15] for the pinhole model. It must be said that our experiments with the pinhole model gave inconsistent results, exhibiting error reduction as reported by Schmalz et al. when using few points to perform the calibration, and showing no effect when using more points. This can be explained by the fact that target shots are mostly frontal to the camera and thus the effect of refraction is mostly radial. Hence, adding a refraction term changes (possibly enlarging) the space of
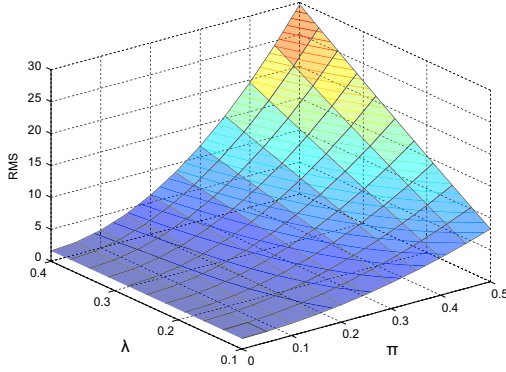
Figure 2. Effect of refraction correction for different values of the refraction parameters.

Figure 3. Manifold interpolation of free rays.

radial functions used for eliminating distortion with the pinhole model. Using more points to perform the calibration constrains the model more, resulting in no additional advantage. Our model, not having an explicit radial distortion term, is not affected by this phenomenon.

To push this finding to the extreme, we placed a 0.75mm glass layer (the front glass of a photo frame) in front of the LCD display in an attempt to produce a much stronger refraction effect. Even in this condition no effect could be measured both in the unconstrained and in the dense pinhole case.

While the estimation of the λ and μ parameters could be incorporated into the calibration process as a third stage performing gradient descent over the parameter λ, the fact that we could not observe any effect moved us to ignore refraction altogether in the experimental evaluation of our approach.

3. Working with the Unconstrained Camera

By alternating the two estimation process we obtain the generalized least squares estimation of both rays and poses, and with that a full calibration of the unconstrained camera model. However, for the unconstrained model to become an effective alternative to the pinhole model several problems must be solved. In particular, if we want to use the model for high precision 3D reconstruction, we need at the very least an effective algorithm for stereo calibration as well as a way to interpolate rays at non-integer coordinates. Potentially we also need a wider set of geometrical and algorithmic tools that are either straightforward of well studied for the pinhole model. In fact, the parametric nature of the pinhole model offers a direct solution the interpolation problem, while there is a ample body of work on how to estimate the motion between two calibrated cameras. As a matter of fact, the pinhole model also allows for useful
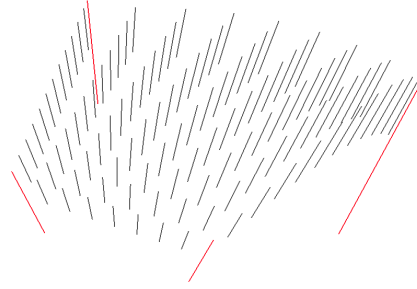
processes like rectification that even conceptually does not have a counterpart in the unconstrained model. However, arguably any measurement or reconstruction process can be reformulated based on only extrinsic calibration and ray interpolation, which incidentally is the minimal requirement to perform triangulation effectively.

For the stereo calibration we adopt a quite crude approach: We take several shots of our active target with both cameras and use the modified ICP algorithm to estimate the poses for each camera. As usual we initialize the poses with the pinhole model to guarantee convergence. With the poses of the first camera at hand, we can construct a set of 3D points $\mathbf{x}_i^s$ in the first camera's coordinate system where ray $\mathbf{r}_i$ intersects the target plane at shot s. Further, given the codes $\mathbf{Co}_i^s$ observed in pixel i at shot s and the pose $\mathbf{\Theta}_s'$ of the target according to the second camera, we can computer the points $\mathbf{y}_i^s$ which represent the points $\mathbf{x}_i^s$ in the coordinate system of the second camera. once we have computed these pairs of points for all camera pixel and shots, we can use Horn's alignment algorithm [7] to estimate the transformation between the two coordinate systems. Clearly this approach has limits, as it does not attempt to reduce the sources of heteroscedasticity, but merely averages over the pose estimations that are by their nature affected by noise, while a more principled approach should pose the problem as global optimization over the observables. However, it is enough to test the feasibility of the unconstrained model and, in conjunction with the unconstrained model, does provide, as will be shown in the experimental section, more precise and repeatable 3D measures than what can be obtained with state-of-the-art approaches based on the pinhole model.

To solve the ray interpolation problem, we generalize bilinear interpolation to the manifold of 3D lines. Under our parametrization, in fact, a line is represented as a point in $\mathbb{R}^6$. However, the normal direction and orthogonal position constraints force the lines to lay in a 4-dimensional manifold. We can generalize (weighted) means over a manifold through the notion of Fréchet means: a point $\mathbf{x}$ residing in
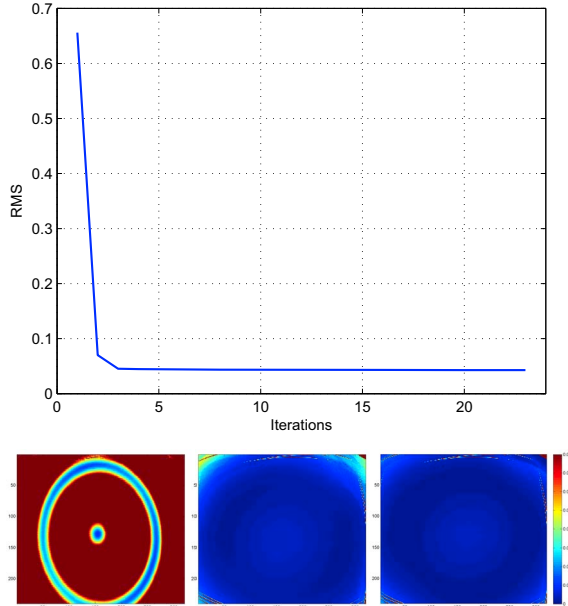
Figure 4. Root mean squared error between expected and observed codes as a function of the number of iterations of the calibration algorithm. The top plot shows the error averaged over all the pixels and poses, while the bottom pictures show for each pixel the error averaged over all the poses at iteration 0 (pinhole model), 2, and 21.

manifold $\mathcal{M}$ is the average of points $\mathbf{x}_i \in \mathcal{M}$ with weights w_i if it solves

$$\underset{\mathbf{x} \in \mathcal{M}}{\arg\min} \sum_i w_i d_{\mathcal{M}}^2(\mathbf{x}, \mathbf{x}_i) \qquad (9)$$

where $d_{\mathcal{M}}$ is the geodesic distance over $\mathcal{M}$. A similar approach has previously been applied to the interpolation of rotations and rigid transformation [16, 11] and, similarly to those approaches, it can be shown that the interpolation is invariant to the frame of reference and interpolates the rays through a minimal path with constant linear and angular velocity. Figure 3 shows that interpolation of four rays (in red) where the weight parameters w_i are taken to be uniformly spaced bi-linear interpolation parameters. Note that, just like in [11], we approximate the Fréchet mean by taking a linear average in $\mathbb{R}^6$ followed by a projection onto the manifold.

4. Experimental Evaluation

In order to evaluate the effectiveness of our calibration approach and the usefulness of the unconstrained model for high-precision vision applications with quasi-pinhole central cameras, we performed several calibration with our method based on a training set of 40 shots of the active target, and tested them on a test set composed of 40 shots.
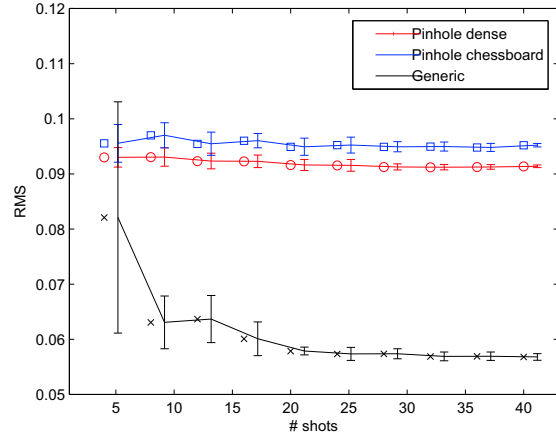
Figure 5. Comparison of the error obtained with the pinhole model calibrated with a chessboard and dense target, and with our calibration approach for the unconstrained model.

For the active target we used an LCD display with a resolution of 1280x1024 pixels, while we used a professional entry level 1 megapixel computer vision camera with variable focal length optics set close to the shortest available length in order to have noticeable, but not extreme distortion. The rays and poses where initialized performing a pinhole calibration using the OpenCV library [3] and adopting a 5th order polynomial model for the radial distortion.

Figure 4 plots the root mean squared error between expected and observed codes as a function of the number of iterations of the calibration algorithm. While the iteration of the calibration procedure was performed on the training set, the computation of the error was performed on the test set by running only the pose estimation without changing the rays. The top figure plots the error averaged over all the pixels and poses, and clearly shows that the estimated model exhibits an order of magnitude lower error than the pinhole model, which is the initialization model ans thus the first entry in the plot. The images in the bottom row display the error for each pixel, averaged over the poses. The leftmost image refers to the initial pinhole model, while the middle and right image refer to the model after 2 and 21 iterations of the calibration procedure. Red pixels represent high error, while blue pixels represent low error. It is immediately apparent that there is still some residual structured error in the pinhole model that the polynomial radial distortion term was not able to correct, resulting in radial waves of high and low error corresponding to areas where the estimated polynomial fits more or less accurately the model. Apart from the radial distortion, however, there is still some error that is not radially symmetric, as can be seen by the different levels of blue and green around the low-error ring. The unconstrained model, on the other hand, not only dramatically reduces the error, but also mitigates the spatial coherency of
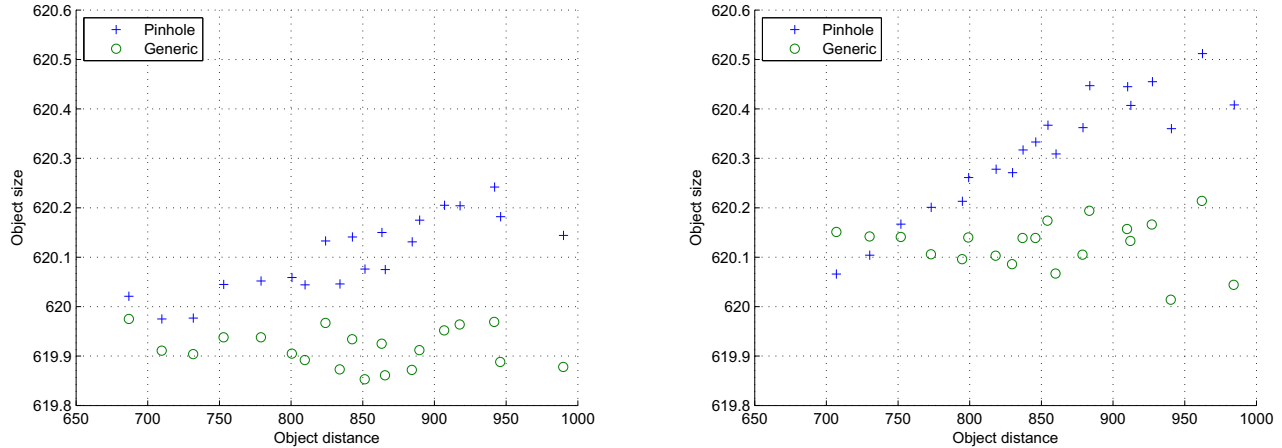
Figure 6. Scatter plot of the measured distances between pairs of points taken in the top part (left) and bottom part (right) of the target. The points where 620 target pixels apart.

the error.

Figure 5 plots the RMS error on the test set obtained with our model as we increase the number of poses in the training set. The plot is compared against the results obtained with a pinhole model calibrated with a standard chessboard pattern and using the same LCD-based active dense target used to calibrate the unconstrained model. We can see that the dense target offers a marginal advantage over the chessboard target for the pinhole model. Note, however, that this is most likely due to the increased precision in the localization of the point offered by the phase shift encoding rather than to the increase in number of points. In fact, no advantage can be seen in adding more shots, since the low dimensional model is already well constrained with a very limited number of poses. The unconstrained model, on the other hand, clearly needs more shots to fully estimate the huge number of free parameters, exhibiting a very large variance when calibrated with few shots, and only really settling down to a precise estimation when at least 20 target points are observed for each pixel. However, while exhibiting large variance, the unconstrained model has a lower average RMS error even with as few as 4 shots, reaching an error when estimated with a sufficient number of shots that is approximately an order of magnitude lower than what can be obtained with the pinhole model.

In order to assess the advantage that the unconstrained model offers over the pinhole one in high precision tasks, we performed a very basic 3D measurement task on a calibrated camera pair. Both cameras where calibrated with the pinhole model (both intrinsic and extrinsic parameters) using the OpenCV library [3], and using the proposed approaches for unconstrained camera and stereo calibration. With the calibration parameters at hand, we triangulated two known points on the calibration target in 20 different

shots in the test set, and computed their distance as a function of their distance from the camera. Figure 6 shows two scatter plots obtained from two different pair of points located in different parts of the target. The target points where, in both cases 620 target pixels apart and the spatial unit of the plot is target pixel width.

From the plot we immediately see that with the pinhole model there is a correlation between depth and measured size, clear indication of an imperfect image formation model, and resulting in a relatively large overall variance of the measure. The unconstrained model, on the other hand, does not exhibit this positional dependency resulting on a much smaller variance in the measurement.

It is worth noting that there is a small bias in the estimated distance. In fact we underestimate the measure of approximately 0.1 pixel out of 620 (about 0.015% or 0.03mm out of a length of approximately 180mm) on the leftmost plot which was drawn from points on the top part of the target, while we overestimate by approximately the same amount on the rightmost plot, taken from points on the bottom part of the target. This can be the result of a non-perfectly rectangular monitor (an error of 0.015% is well within the construction tolerances) and can be seen in the pinhole model as well.

5. Discussion

In this paper we investigated the use of an unconstrained camera model to calibrate central quasi-pinhole cameras for high precision measurement and reconstruction tasks, and provided an effective approach to perform the calibration. The basic ingredient for the calibration process is the use of a dense target which allows us to attain a favorable parameters to measurements ratio, guaranteeing a stable calibration with a limited number of shots, and thus rendering the pro-
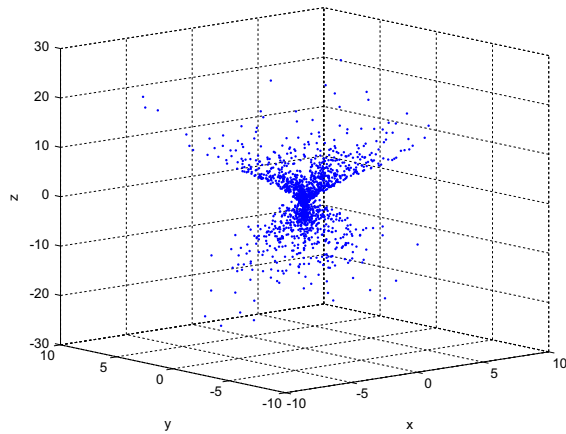
Figure 7. Spatial distribution of the local pinholes, i.e., of the closet point to the rays in a local neighborhood. In a pinhole model all the points would coincide. For display purposes the points are sub-sampled.

cess operationally not much more complex than standard pinhole calibration with sparse passive targets.

The resulting model can successfully eliminate the spatial coherence of the error, resulting in more precise and repeatable measures than what is achieved with the pinhole model. In fact there is a clear indication that the estimated models are substantially different from a radially undistorted pinhole model. To see this difference, we can visualize how far the rays are from passing through a single point. To this effect, for each pixel and its 4-neighborhood, we define a local pinhole as the point that has minimal sum of squared distances from the five rays.

Figure 7 shows the spatial distribution of these local pinholes, sub-sampled for display purposes. It is clear that the distribution is quite different from a random spreading of coincident points, exhibiting a very distinctive spatial coherency, as teh points lay quite tightly on a clear manifold linked with the evolute of the calibrated rays.

Clearly, this is only a preliminary analysis, and much work still needs to be done before the unconstrained model can substitute effectively the pinhole model. In particular, more principled approaches to stereo calibration are needed, as well as alternatives for those algorithms that rely on geometrical processes that do not have a counterpart on the unconstrained model. However, we feel that there is much to be gained by moving towards non-parametric camera models, espetially in high precision 3D vision tasks.

References

[1] D. G. Aliaga. Accurate catadioptric calibration for real-time pose estimation of room-size environments. In *International Conference of Computer Vision*, pages 127–134, 2001. 1

[2] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992. 4

[3] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, Inc., 1st edition, October 2008. 6, 7

[4] F. Devernay and O. Faugeras. Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured enviroments. *Mach. Vision Appl.*, 13(1):14–24, Aug. 2001. 1

[5] K. Gremban, C. Thorpe, and T. Kanade. Geometric camera calibration using systems of linear equations. In *Proceedings of IEEE Conference on Robotics and Automation (ICRA '88)*, volume 1, pages 562 – 567, April 1988. 1

[6] M. D. Grossberg and S. K. Nayar. A general imaging model and a method for finding its parameters. In *International Conference of Computer Vision*, pages 108–115, 2001. 1

[7] B. K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. of the Optical Society of America. A*, 4(4):629–642, Apr 1987. 4, 5

[8] C. Hughes, P. Denny, E. Jones, and M. Glavin. Accuracy of fish-eye lens models. *Appl. Opt.*, 49(17):3338–3347, Jun 2010. 1

[9] J. Kannala and S. S. Brandt. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(8):1335–1340, Aug. 2006. 1

[10] T. Kariya and H. Kurata. *Generalized Least Squares*. Wiley, 2004. 3

[11] L. Kavan, S. Collins, C. O'Sullivan, and J. Žára. Dual quaternions for rigid transformation blending. Technical Report TCD-CS-2006-46, Trinity College Dublin, 2006. 6

[12] E. Lilienblum and B. Michaelis. Optical 3d surface reconstruction by a multi-period phase shift method. *JCP*, 2(2):73–83, 2007. 2

[13] O. Morel, R. Seulin, and D. Fofi. Catadioptric camera calibration by polarization imaging. In *Pattern Recognition and Image Analysis*, volume 4478 of *Lecture Notes in Computer Science*, pages 396–403. Springer, 2007. 1

[14] S. Ramalingam, P. Sturm, and S. K. Lodha. Towards complete generic camera calibration. In *Proc. IEEE Computer Vision and Pattern Recognition*, pages 1093–1098, 2005. 2

[15] C. Schmalz, F. Forster, and E. Angelopoulou. Camera calibration: active versus passive targets. *Optical Engineering*, 50(11), 2011. 4

[16] K. Shoemake. Animating rotation with quaternion curves. In *Proc. of the 12th annual conference on Computer graphics and interactive techniques*, pages 245–254, 1985. 6

[17] P. Sturm and S. Ramalingam. A generic calibration concept: Theory and algorithms. Technical Report 5058, INRIA, dec 2003. 2

[18] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In *Proc. European Conference on Computer Vision*, volume 2, pages 1–13. Springer, May 2004. 1

[19] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987. 1