Automatic Scene Structure and Camera Motion using a Catadioptric System

Maxime Lhuillier

Automatic Scene Structure and Camera Motion using a Catadioptric System

Maxime Lhuillier
LASMEA-UMR 6602, UBP/CNRS,
24 avenue des Landais,
63177 Aubière Cedex, France.
Mail: Maxime.Lhuillier@univ-bpclermont.fr
Tel: +33(0)4 73 40 75 93
Fax: +33(0)4 73 40 72 62

The reference of this paper is: Maxime Lhuillier, Automatic Scene Structure and Camera Motion using a Catadioptric System, Computer Vision and Image Understanding, 109(2):186-203, 2008.

Automatic Scene Structure and Camera Motion using a Catadioptric System ⋆

Maxime Lhuillier [a,*]

[a] *LASMEA-UMR 6602, UBP/CNRS*
24 avenue des Landais, 63177 Aubière Cedex, France.

Abstract

Fully automatic methods are presented for the estimation of scene structure and camera motion from an image sequence acquired by a catadioptric system. The first contribution is the design of bundle adjustments for both central and non-central models, by taking care of the smoothness of the minimized error functions. The second contribution is an extensive experimental study for long sequences of catadioptric images in a context useful for applications: a hand-held and equiangular camera moving on the ground. An equiangular camera is non-central and provides uniform resolution in the image radial direction. Many experiments dealing with robustness, accuracy, uncertainty, comparisons between both central and non-central models, and piecewise planar 3D modeling are provided.

Key words: Central and Non-Central Catadioptric Cameras, Structure from Motion, Bundle Adjustment, Vision System

⋆ Expanded version of a paper accepted at the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS'05), Beijing, October 2005

* Corresponding author
Email address: Maxime.Lhuillier@univ-bpclermont.fr (Maxime Lhuillier).
URL: maxime.lhuillier.free.fr (Maxime Lhuillier).

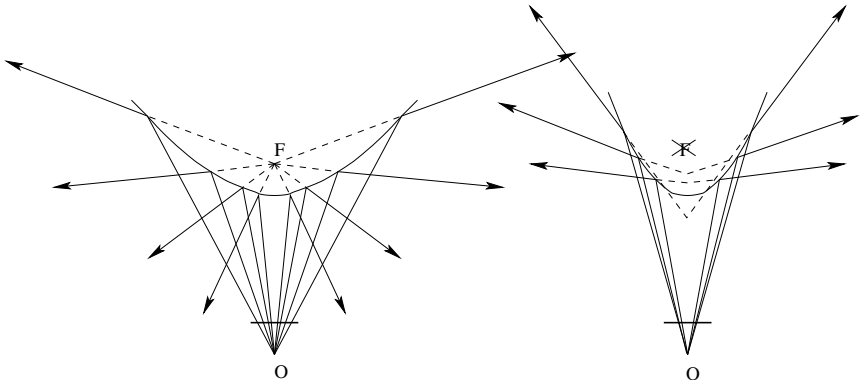

Fig. 1. Two catadioptric systems obtained with a convex mirror in front of a perspective camera (with projection center O). Left: a central system (all extended rays go through a single point F, called the center). Right: a non-central system.

1 Introduction

The automatic estimation of scene structure and camera motion from image sequence acquired by a perspective camera (whether calibrated or not) taken by hand has been a very active topic [11,6] during recent decades, and many successful systems now exist [3,22,21,15]. This paper (an extended version of [16]) focuses on this estimation using a catadioptric system/camera. Such a system has a wide field of view thanks to a convex mirror mounted in front of a perspective camera as shown in Figure 1. A catadioptric system may be "central" if all back-projected rays go through a single point in 3D called the "center". An overview of the paper and comparisons with previous works are given in Sections 1.1 and 1.2, respectively. Section 1.3 provides a brief summary on the known desirable conditions for errors minimized by bundle adjustments and discusses errors introduced in previous works.

1.1 Overview

First, central and non-central catadioptric camera models are described in Section 2. The non-central model explicitly involves the ray reflection onto the mirror surface, while the central model directly defines a mapping from 3D to 2D. These models require neither a conic as mirror profile nor the perspective camera center at one of the conic focal points.

Second, Section 3 presents bundle adjustments for a catadioptric camera in a general context: field of view greater than 180°, scene with distant 3D points, approximate calibration knowledge and image noise. Although bundle adjustment is a well known iterative method, we spend some time in Sections 3.1 and 3.2 going over desirable smoothness conditions on the minimized errors

to facilitate the convergence. These conditions are not fully satisfied by the standard 2D reprojection error of a catadioptric camera. With this in mind, we propose bundle adjustment methods minimizing 2D image errors (measured in pixels) in Section 3.3 and angular errors (measured in radians) in Section 3.4.

Third, Section 4 describes the automatic estimation method. The geometry of the image sequence is estimated for the central model. Then, a second estimation using the non-central model is obtained. In both cases, an approximate and given calibration is refined with the 3D.

Fourth, Section 5 presents an extensive experimental study in a context that we expect to be representative (although not trivial) and useful for applications: a hand-held and equiangular catadioptric camera moving on the ground. An equiangular camera is non-central and provides uniform resolution in the image radial direction. We propose experiments about robustness, accuracy and uncertainty estimation, performance comparisons between central and non-central models, and piecewise planar 3D modeling from the reconstructed points.

The two main contributions of this paper are the bundle adjustments with image and angle errors in Section 3, and the experimental study in Section 5.

1.2 Comparisons with Previous Works

The most related work is that of Micusik and Pajdla [20]. It is a two step method like ours: (1) estimate the sequence geometry using a central model for the camera and (2) upgrade the sequence geometry using a non-central model enforcing the mirror knowledge. The central model is similar to ours and is more general than the Geyer-Daniilidis model [7] which requires a conic as mirror profile and the perspective camera center at one of the conic focal points. A general central model is more adequate to approximate a (general) non-central catadioptric camera. The differences are the following. The non-central model proposed in [20] is formalized from 2D to 3D since the authors minimize a 3D error for bundle adjustment. Our 3D to 2D formalization is more adequate for bundle adjustment minimizing a 2D image error (measured in pixels). Comparisons between 3D and 2D errors are made in Section 1.3. Furthermore, the work [20] emphasizes auto-calibration in a general context for two views only. This is not our focus: we assume that an approximate calibration is given and we emphasize (1) the definition of errors with high smoothness to be minimized by multi-view bundle adjustment and (2) extensive experiments on long sequences. From a technical viewpoint, we use the 7-point algorithm [11] to estimate fundamental matrix instead of the 9 and 15-point methods [20] involving polynomial eigenvalue problems. Both papers

provide experiments with a non-central catadioptric camera: a (roughly) orthographic camera with a parabolic mirror [20], and a (roughly) equiangular catadioptric camera in our case. The mirror profile of our equiangular camera is not a conic.

A different problem [2] is the real-time pose estimation of a parabolic catadioptric camera for planar motions in a room-size environment, using a few known 3D beacons. The ideal model is central: a parabolic mirror in front of an orthographic camera. The real model is non-central: a parabolic mirror in front of a perspective camera. In this context, the pose accuracy is improved by the non-central model. In our context, experiments show that the advantages of the non-central model are not so convincing.

The geometry of a catadioptric sequence without accurate calibration has been estimated in other works, but they involve two views and use central camera models. Calibration and successive essential matrices are estimated for a parabolic catadioptric camera from tracked points in the image sequence [13]. Calibration initialization is given by the approximate field of view angle. Other work on the same sensor [8] introduces the parabolic fundamental matrix, which defines a bilinear constraint between two matched points represented in an adequate space, and shows that calibration and essential matrix recovery are possible from this matrix.

1.3 Which Errors for Bundle Adjustment ?

Bundle adjustment (BA) is the standard process to obtain an accurate estimate of the geometry (all cameras and 3D points) by minimizing a sum of squares of non-linear errors. A survey about bundle adjustment [26] (also refer to [11,19]) describes the desiderata for these errors in detail.

First, errors should be smooth enough in order to facilitate the convergence of BA. C^2 continuity is recommended to reach quadratic final convergence using the Levenberg-Marquardt method involved in BA [19,26]. This continuity includes the case of a 3D point at the infinity plane thanks to the homogeneous coordinates [26]. It is easy to satisfy this condition for a conventional (perspective) camera, but it is not for a catadioptric camera. Indeed, the definition of errors with high smoothness is not straightforward in our context since the catadioptric projection function is always discontinuous at the infinite plane.

Second, errors should be well chosen in order to improve the geometry estimation in a statistical sense [26]. Minimized errors are usually noisy measurements of physical quantities and are modeled by realizations of random vectors. We obtain a Maximum Likelihood Estimation (MLE) of the geometry with the common statistical model [11]: errors obey zero-mean isotropic Gaus-

sian probability distributions which are independent and identical (outliers are removed before the BA step). The minimization of 2D reprojection errors has been used for many decades, whereas 3D error minimizations [12,17,20] provide biased results as mentioned in [12,17]. The reason is the following: the common Gaussian model is not tenable for 3D errors (distances between back-projected rays and 3D reconstructed points) because the true probability distribution of 3D error depends too much on the distance between cameras and the point. However, the bias is limited in experiments [12,20] since 3D points in indoor scenes have distances of the same magnitude order. More details on the 3D error in papers [12,17,20] are given below.

The 3D error is used for the point estimation by intersection of rays [12] (page 181). Let $r(\lambda_k) = \mathbf{t}_k + \lambda_k \mathbf{v}_k$ be the parametrization of the k^{th} ray with $\mathbf{t}_k$ and $\mathbf{v}_k$ the starting point and normalized direction of the ray. The 3D point $\mathbf{p}$ closest to all of the rays minimizes $\sum_k ||\mathbf{e}_k||^2$ with the error $\mathbf{e}_k = \mathbf{p} - (\mathbf{t}_k + \lambda_k \mathbf{v}_k)$. No iterative method is needed since a closed form expression of $\mathbf{p}$ exists. The authors mention that they obtain a more "optimal" result by minimizing $\sum_k \frac{1}{\lambda_k^2} ||\mathbf{e}_k||^2$. They introduce the weights $\frac{1}{\lambda_k^2}$ and argue that the uncertainty in point location grows linearly with λ_k. In other words, the common Gaussian model above is more tenable for $\frac{\mathbf{e}_k}{\lambda_k}$ than $\mathbf{e}_k$. Furthermore, we note that $||\frac{\mathbf{e}_k}{\lambda_k}|| = ||\frac{\mathbf{p}-\mathbf{t}_k}{\lambda_k} - \mathbf{v}_k||$ at the solution is similar to a distance between two close points onto the unit sphere: it is an angle between two directions. Section 3.4 describes an angle error which is C^2 continuous, even at the infinite plane. The errors of [12] are discontinuous at the infinite plane.

The 3D error is used for the estimation of the camera pose [17]: find the pose given many 3D points in the world coordinate system and the corresponding ray directions in the camera coordinate system. The method is globally convergent, but the authors mention and experiment that the pose solution more heavily weights the points that are farther away from the camera.

Last, the 3D error is used for bundle adjustment with a non-central catadioptric camera [20]. In this context, the 3D error is preferred to the 2D reprojection error which is more time consuming. The reason is that the catadioptric projection function has no closed form for a general mirror. The bundle adjustment is designed for two views only. The cost function $\sum_j e_j^2$ is minimized with error $e_j = (\mathbf{t}_j^0 - \mathbf{t}_j^1).\frac{\mathbf{v}_j^0 \times \mathbf{v}_j^1}{||\mathbf{v}_j^0 \times \mathbf{v}_j^1||}$. In this expression, the j^{th} reconstructed point is in the center of the shortest transversal of two back-projected rays defined by $(\mathbf{t}_j^0, \mathbf{v}_j^0)$ and $(\mathbf{t}_j^1, \mathbf{v}_j^1)$ ($\mathbf{t}_j^i$ and $\mathbf{v}_j^i$ are the starting point and normalized direction of the back-projected ray by the i^{th} camera). We see that the point parameters are eliminated. Drawbacks are the 2-view use, and the bias mentioned by previous authors for point [12] and camera pose [17] estimations.

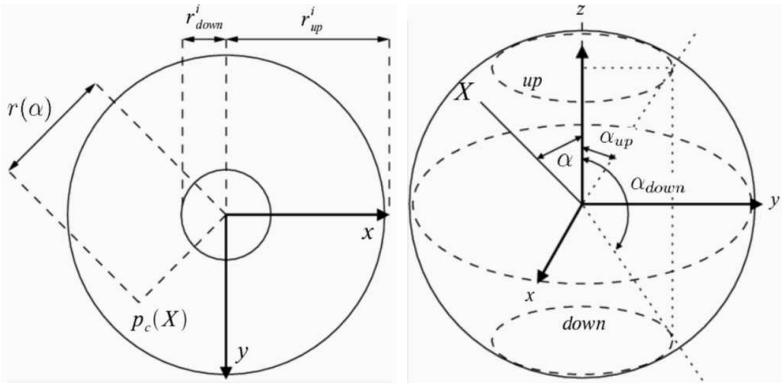

Fig. 2. Left: two concentric circles of radii r^i_{up}, r^i_{down} in the image. Right: angles $\alpha_{up}, \alpha_{down}, \alpha$ from the z-axis define the field of view for a point $\mathbf{X}$ by $\alpha_{up} \leq \alpha(\mathbf{X}) \leq \alpha_{down}$. The circle "up" (respectively, "down") of rays on the right is projected by the central model on the large (respectively, small) circle on the left.

2 Catadioptric Camera Models

Figure 5 shows a catadioptric camera we model: the image projection of the scene is between a large circle and a small circle. Sections 2.1 and 2.2 present the projection functions of the central and non-central camera models for a finite 3D point, respectively. Section 2.3 introduces the definition of antipodal projections. The case of a point in the infinite plane is detailed in Section 2.4.

2.1 Central Model

The central catadioptric model is defined by its orientation $\mathtt{R}$ (a rotation), the center $\mathbf{t} \in \Re^3$ ($\Re$ is the real numbers), both expressed in the world coordinate system, and a projection function C from $\Re^3$ to $\Re^2$ defined below. If $\mathbf{X}$ is the homogeneous world coordinates of a finite 3D point such that the last coordinate is positive, the projection of $\mathbf{X}$ in the image is defined by

$$p_c(\mathbf{X}) = C(\mathbf{d}), \quad \mathbf{d} = \mathtt{R}^\top [\mathtt{I}_3 | -\mathbf{t}]\mathbf{X}. \tag{1}$$

Vector $\mathbf{d}$ is the direction of the ray from $\mathbf{t}$ to $\mathbf{X}$ in the camera coordinate system. The sign constraint for $\mathbf{X}$ is necessary to choose a ray (half line) among the opposite rays defined by directions $\mathbf{d}$ and $-\mathbf{d}$. Indeed, these two rays are projected on two different points in the omnidirectional image if the field of view is greater than 180°.

The central model also has a symmetry around the z-axis of the camera coordinate system: the omnidirectional image is between two concentric circles of radii r^i_{up} and r^i_{down}, and C is such that there is a positive and decreasing

function r such that

$$C(x,y,z) = r(\alpha(x,y,z))\frac{1}{\sqrt{x^2+y^2}}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{2}$$

Let $\alpha = \alpha(x,y,z) = \alpha(\mathbf{d})$ be the angle between the z-axis and the ray direction $\mathbf{d}$. The two angles α_{up} and α_{down} which define the field of view are such that $\alpha_{up} \leq \alpha \leq \alpha_{down}$ and $r^i_{down} \leq r(\alpha) \leq r^i_{up}$ (see Figure 2). We assume that r and p_c are C^2 continuous functions.

2.2 Non-Central Model

The non-central catadioptric model is defined by the orientation $\mathtt{R}$ (a rotation) and the location $\mathbf{t} \in \Re^3$ of the mirror coordinate system, both expressed in the world coordinate system, the intrinsic parameters $\mathtt{K}_p$, orientation $\mathtt{R}_p$ and location $\mathbf{t}_p$ of the perspective camera expressed in the mirror coordinate system, and the (known) mirror surface.

The mirror surface has a symmetry around the z-axis of the mirror coordinate system. $\mathbf{t}$ is at the mirror apex. Let $M^+(\mathbf{A},\mathbf{B})$ be the function which gives the reflection point on the mirror for the ray which goes across points $\mathbf{A}$ and $\mathbf{B}$ (the mirror profile is such that there is at most one reflection point). Point $\mathbf{A}$ is in the incident part and $\mathbf{B}$ is in the reflected part of the ray, or vice versa. We have $\mathbf{A},\mathbf{B},M^+(\mathbf{A},\mathbf{B}) \in \Re^3$, all expressed in the mirror coordinate system. This function does not have a closed form, and its calculation from the mirror surface is presented in Appendix D.

Let $\mathbf{X}$ be the homogeneous world coordinates of a finite 3D point and

$$\pi_2((\begin{array}{ccc} x & y & z \end{array})^\top) = (\begin{array}{cc} \frac{x}{z} & \frac{y}{z} \end{array})^\top, \quad \pi_3((\begin{array}{ccc} x & y & z & t \end{array})^\top) = (\begin{array}{ccc} \frac{x}{t} & \frac{y}{t} & \frac{z}{t} \end{array})^\top. \tag{3}$$

Since $\mathtt{R}^\top(\pi_3(\mathbf{X}) - \mathbf{t})$ and $\mathtt{K}_p\mathtt{R}_p^\top[\mathtt{I}_3|-\mathbf{t}_p]$ are respectively the $\mathbf{X}$ coordinates and the perspective camera matrix in the mirror coordinate system, the projection of $\mathbf{X}$ in the image is defined by

$$p^+_{nc}(\mathbf{X}) = \pi_2(\mathtt{K}_p\mathtt{R}_p^\top(M^+(\mathbf{t}_p, \mathtt{R}^\top(\pi_3(\mathbf{X}) - \mathbf{t})) - \mathbf{t}_p)). \tag{4}$$

Figure 3 shows a cross section of the mirror and pinhole camera, the points $\mathbf{t}_p, \mathbf{X}, p^+_{nc}(\mathbf{X}), M^+(\mathbf{t}_p,\mathbf{X})$ with $\mathtt{R} = \mathtt{I}_3$ and $\mathbf{t} = 0$. Figure 3 also defines a kind of antipodal mirror point $M^-(\mathbf{t}_p,\mathbf{X})$ for $M^+(\mathbf{t}_p,\mathbf{X})$ and its projection $p^-_{nc}(\mathbf{X})$ which will be useful later. Obviously, we have

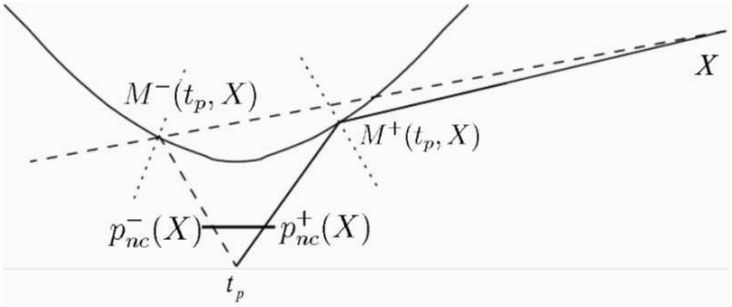

Fig. 3. The ray which goes across finite point $\mathbf{X}$ and pinhole camera center $\mathbf{t}_p$ is reflected on the mirror at point $M^+(\mathbf{t}_p, \mathbf{X})$, and projected by the pinhole camera to $p^+_{nc}(\mathbf{X})$. We also define the point $M^-(\mathbf{t}_p, \mathbf{X})$ on the mirror, projected by the pinhole camera to $p^-_{nc}(\mathbf{X})$. The only difference between $M^-(\mathbf{t}_p, \mathbf{X})$ and $M^+(\mathbf{t}_p, \mathbf{X})$ is the following: the reflected ray from $M^+(\mathbf{t}_p, \mathbf{X})$ (respectively, $M^-(\mathbf{t}_p, \mathbf{X})$) towards $\mathbf{X}$ is pointing outside (respectively, inside) the mirror. In general, $\mathbf{X}, M^+(\mathbf{t}_p, \mathbf{X})$ and $M^-(\mathbf{t}_p, \mathbf{X})$ are not collinear points.

$$p^-_{nc}(\mathbf{X}) = \pi_2(\mathtt{K}_p \mathtt{R}_p^\top (M^-(\mathbf{t}_p, \mathtt{R}^\top(\pi_3(\mathbf{X}) - \mathbf{t})) - \mathbf{t}_p)). \tag{5}$$

We note that the parametrization $(\mathtt{R}_p, \mathbf{t}_p, \mathtt{R})$ is not minimal because of the mirror symmetry around the z-axis: the expressions of $p^+_{nc}(\mathbf{X})$ and $p^-_{nc}(\mathbf{X})$ are unchanged by replacing $(\mathtt{R}_p, \mathbf{t}_p, \mathtt{R})$ by $(\mathtt{R}_z\mathtt{R}_p, \mathtt{R}_z\mathbf{t}_p, \mathtt{R}\mathtt{R}_z^\top)$ and applying the mirror symmetry relations

$$\mathtt{R}_z M^+(\mathbf{A}, \mathbf{B}) = M^+(\mathtt{R}_z\mathbf{A}, \mathtt{R}_z\mathbf{B}), \quad \mathtt{R}_z M^-(\mathbf{A}, \mathbf{B}) = M^-(\mathtt{R}_z\mathbf{A}, \mathtt{R}_z\mathbf{B})$$

for any rotation $\mathtt{R}_z$ around the z-axis. Last, we assume in this paper that the mirror profile is such that M^+ and M^- are C^2 continuous functions. The resulting $p^+_{nc}(\mathbf{X})$ and $p^-_{nc}(\mathbf{X})$ are C^2 continuous.

2.3 Antipodal Projections

The antipodal projections of $\mathbf{X}$ are defined by $\{p_c(\mathbf{X}), p_c(-\mathbf{X})\}$ in the central case, and by $\{p^+_{nc}(\mathbf{X}), p^-_{nc}(\mathbf{X})\}$ in the non-central case.

Assume that both antipodal projections exist. In the central case, the antipodal projections are $C(\mathbf{d})$ and $C(-\mathbf{d})$ with $\mathbf{d} \in \Re^3$. They are separated by the small circle. In the non-central case, Figure 3 shows that the small circle also separate the two antipodal projections. In practice, the distance between the two antipodal points is greater than the diameter of the small circle (hundreds of pixels in the experiments).

Let $\mathbf{X}_\infty = (\,x \quad y \quad z \quad 0\,)^\top$ be a point in the infinite plane. We consider the finite point $X(t) = (\,\frac{x}{t} \quad \frac{y}{t} \quad \frac{z}{t} \quad 1\,)^\top$ such that $t \in \Re$ converges to 0. The limit of $X(t)$ is the infinite point $\mathbf{X}_\infty$ in the projective space, and two different limits of $p_c(X(t))$ or $p_{nc}^+(X(t))$ are possible in the catadioptric image: one for each sign of t. These limits (if they exist) define the projection(s) of $\mathbf{X}_\infty$ by the catadioptric camera.

The main subject of this part is to show that the limit of each antipodal projection (defined in Section 2.3) of $X(t)$ is a projection of $\mathbf{X}_\infty$.

Obviously, the limits of the first antipodal projections

$$\lim_{\begin{cases} t \to 0 \\ t > 0 \end{cases}} p_c(X(t)), \qquad \lim_{\begin{cases} t \to 0 \\ t > 0 \end{cases}} p_{nc}^+(X(t))$$

are projections of $\mathbf{X}_\infty$ for the central and non-central models, respectively.

In the central case, we introduce $d(t) = \frac{\mathtt{R}^\top[\mathtt{I}_3|-\mathbf{t}]X(t)}{||\mathtt{R}^\top[\mathtt{I}_3|-\mathbf{t}]X(t)||}$ which verifies

$$p_c(X(t) = C(d(t)), \quad p_c(-X(t) = C(-d(t)), \quad \lim_{\begin{cases} t \to 0 \\ t < 0 \end{cases}} d(t) = -\lim_{\begin{cases} t \to 0 \\ t > 0 \end{cases}} d(t).$$

Thanks to the continuity of C onto the unit sphere,

$$\lim_{\begin{cases} t \to 0 \\ t > 0 \end{cases}} p_c(-X(t)) = C(-\lim_{\begin{cases} t \to 0 \\ t > 0 \end{cases}} d(t)) = C(\lim_{\begin{cases} t \to 0 \\ t < 0 \end{cases}} d(t)) = \lim_{\begin{cases} t \to 0 \\ t < 0 \end{cases}} p_c(X(t)).$$

Thus, the limit of the second antipodal projection (on the left) is the second projection of $\mathbf{X}_\infty$ (on the right).

In the non-central case, a proof sketch is given. Figure 4 is obtained from Figure 3 for a finite point $\mathbf{X} = X(t)$ which converges to the right toward $\mathbf{X}_\infty$ with $t > 0$. The points $M^+(\mathbf{t}_p, \mathbf{X})$, $p_{nc}^+(\mathbf{X})$, $M^-(\mathbf{t}_p, \mathbf{X})$, $p_{nc}^-(\mathbf{X})$ converge to $M^+(\mathbf{t}_p, \mathbf{X}_\infty^+)$, $p_{nc}^+(\mathbf{X}_\infty^+)$, $M^-(\mathbf{t}_p, \mathbf{X}_\infty^+)$, $p_{nc}^-(\mathbf{X}_\infty^+)$, respectively. We see in Figure 4 that the back projected ray of pixel $p_{nc}^-(\mathbf{X}_\infty^+)$ is reflected by the mirror at $M^-(\mathbf{t}_p, \mathbf{X}_\infty^+)$ with the direction given by $\mathbf{X}_\infty$ toward the left. Thus, the limit $p_{nc}^-(\mathbf{X}_\infty^+)$ of the second antipodal projection $p_{nc}^-(\mathbf{X}(t))$ is one of the two projections of $\mathbf{X}_\infty$. Furthermore, it is not $p_{nc}^+(\mathbf{X}_\infty^+)$, the first projection of $\mathbf{X}_\infty$.

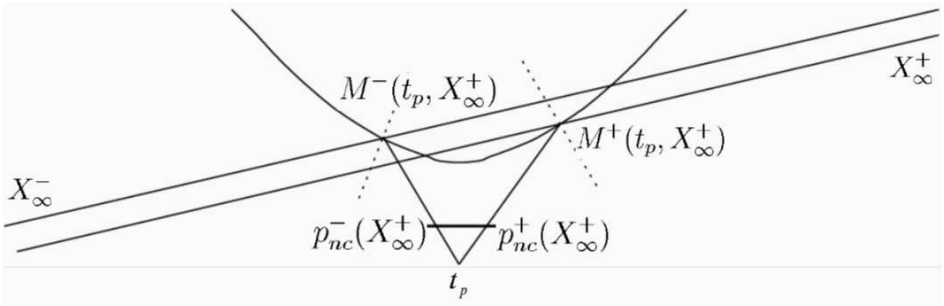

Fig. 4. This figure is obtained if the finite point $\mathbf{X}$ of Figure 3 converges toward the right to the infinite point $\mathbf{X}^+_\infty$. The two lines which go across $M^+(\mathbf{t}_p, \mathbf{X}^+_\infty)$ and $M^-(\mathbf{t}_p, \mathbf{X}^+_\infty)$ have the same direction $\mathbf{X}^+_\infty$.

3 Bundle Adjustments for a Catadioptric Camera

Bundle adjustment [11] or BA is the standard process to obtain an accurate estimate of the sequence geometry. It iteratively adjusts the parameters of all cameras and all 3D points in order to minimize a cost function. This function is a sum of squares of non-linear errors

$$E = \sum_{i,j} ||\mathbf{e}^i_j||^2 \tag{6}$$

with $\mathbf{e}^i_j$ the error corresponding to the i^{th} camera and the j^{th} point. $||\mathbf{x}||$ is the Euclidean norm of a vector $\mathbf{x}$. The standard 2D reprojection error is $\mathbf{e}^i_j = p^i(\mathbf{X}_j) - \mathbf{u}^i_j$ (measured in pixels), where p^i is the projection function by the i^{th} camera, $\mathbf{X}_j$ is the j^{th} 3D point in the world coordinate system, and $\mathbf{u}^i_j$ the point detected in the i^{th} image which corresponds to the j^{th} point. Sections 3.1 and 3.2 describe smoothness conditions on $\mathbf{e}^i_j$. Then, we introduce the image error in Section 3.3 and the angular error in Section 3.4.

3.1 Smoothness Conditions

The most commonly used minimization tool for geometry refinements is the Levenberg-Marquardt method (LM), which combines the Gauss-Newton algorithm and the descent of gradient [19]. Thus, a summary of LM is useful to justify conditions on $\mathbf{e}^i_j$. LM requires a starting point and produces a series of vectors which converges to a local minimizer of the cost function. This local minimizer is the right one if the starting point is close enough. Let $\mathbf{p}_n$ be the current value of the geometry parameters. The quadratic Taylor expansion of the cost function defined by Equation 6 is

$$E(\mathbf{p}_n + \mathbf{\Delta}) \approx E(\mathbf{p}_n) + \mathbf{\Delta}^\top \mathbf{E}'_n + \frac{1}{2}\mathbf{\Delta}^\top \mathtt{E}''_n \mathbf{\Delta}$$

with $\mathbf{E}_n^{'}$ and $\mathtt{E}_n^{''}$ the gradient and Hessian of E at $\mathbf{p}_n$. One LM iteration is $\mathbf{p}_{n+1} = \mathbf{p}_n + \mathbf{\Delta}_n$ where $\mathbf{\Delta}_n$ satisfies $(\mathtt{E}_n^{''} + \lambda_n \mathtt{I})\mathbf{\Delta}_n = -\mathbf{E}_n^{'}$, and λ_n is a well chosen parameter evolving during the convergence such that $E(\mathbf{p}_n + \mathbf{\Delta}_n) < E(\mathbf{p}_n)$. $\mathtt{I}$ is the identity matrix (or a diagonal matrix obtained by the diagonal coefficients of $\mathtt{E}_n^{''}$ [23]). If λ_n is "small", the LM behavior is Newton like: we get quadratic convergence in the immediate neighborhood of the solution if E is C^2 continuous and $\mathtt{E}_n^{''}$ is definite positive. Elsewhere, a gradient descent like behavior is preferred with a "large" λ_n to guarantee a decreasing cost function.

In practice, the Gauss-Newton approximation is always used for the Hessian $\mathtt{E}_n^{''}$ of least squares:

$$\frac{1}{2}\frac{\partial^2(||\mathbf{e}_j^i||^2)}{\partial p_0 \partial p_1} = {\mathbf{e}_j^i}^\top \frac{\partial^2 \mathbf{e}_j^i}{\partial p_0 \partial p_1} + \frac{\partial \mathbf{e}_j^i}{\partial p_0}^\top \frac{\partial \mathbf{e}_j^i}{\partial p_1} \approx \frac{\partial \mathbf{e}_j^i}{\partial p_0}^\top \frac{\partial \mathbf{e}_j^i}{\partial p_1} \tag{7}$$

where p_0, p_1 are any parameters of the geometry. This approximation is acceptable if $||\mathbf{e}_j^i||$ is small and $||\frac{\partial^2 \mathbf{e}_j^i}{\partial p_0 \partial p_1}||$ is bounded.

This summary about LM gives many conditions for an ideal error. First, the C^1 continuity of $\mathbf{e}_j^i$ is sufficient for the evaluation of $\mathbf{e}_j^i$ Jacobian (Equation 7). Second, the C^2 continuity of $\mathbf{e}_j^i$ is even better to reach the quadratic final convergence, with the additional condition: $\mathbf{e}_j^i = 0$ if the i^{th} camera and the j^{th} point are in exact agreement. Third, the Jacobians of $\mathbf{e}_j^i$ should not all be 0 if $\mathbf{e}_j^i = 0$. Otherwise, the $\mathtt{E}_n^{''}$ approximation by Equation 7 converges to zero at the final convergence of LM and the quadratic convergence is lost.

3.2 3D Point Parametrization and Crossing the Infinite Plane

In this part, we spend some time explaining why we should also pay attention to the parametrization of the 3D point $\mathbf{X}_j$ involved in $\mathbf{e}_j^i$ in the neighborhood of the infinite plane. The case of $\mathbf{e}_j^i$ defined by the standard 2D reprojection error is studied (a similar discussion on this topic is also given in [26]).

Assume that the $\mathbf{X}_j$ parametrization is (x, y, z). Points onto the infinite plane are ignored by this parametrization, although they may have a projection by the camera. Furthermore, major changes of x, y and z are needed to modify $||\mathbf{e}_j^i||$ if the 3D point $\mathbf{X}_j$ is distant from the i^{th} camera. Now, the affine parametrization (x, y, z) is replaced by the homogeneous projective parametrization (x, y, z, t). This parametrization allows $\mathbf{X}_j$ to go across (and onto) the infinite plane with small variations of t and x, y, z. After a crossing of the infinite plane (if the sign of t changes during an LM iteration), a point in front of the camera goes behind the camera and vice versa. A crossing of the

infinite plane is just a kind of "faster way" to connect two distant 3D points of a camera. We note that $\mathbf{e}_j^i$ for a perspective camera is C^2 continuous with the 3D point parametrization (x, y, z, t), even if $\mathbf{X}_j$ is in the infinite plane.

As mentioned in [26], such a crossing of the infinite plane by a point is sometimes necessary for the bundle adjustment convergence due to noise or approximate calibration. An approximate calibration or noise may reconstruct many distant points (by ray intersection) behind the camera although they should be in front of it, and the only way to escape from this configuration using an LM iteration may be by crossing the infinite plane. Furthermore, such transitory configurations help with LM convergence.

Now, we focus on the case of any catadioptric camera with $\mathbf{e}_j^i$ defined by the standard 2D reprojection error and the parametrization (x, y, z, t). This was not done in [26]. A 3D point in the infinite plane is also visible, and it may have two projections by the camera if the field of view is greater than 180°. In this case, Section 2.4 shows that the catadioptric projection functions p_c and p_{nc}^+ are not continuous at the infinite plane since two different limits are obtained. Thus, $\mathbf{e}_j^i$ is not a continuous function at the infinite plane (if $t = 0$). This error is not ideal since we do not have the C^2 continuity for the crossing of the infinite plane. Even worse, the C^0 discontinuity of $\mathbf{e}_j^i$ is very great since the two projection limits (if t converge to 0) are two antipodal points separated by the small circle of the catadioptric image. If $\mathbf{X}_j$ goes from the right side (where $||\mathbf{e}_j^i||$ is less than 2 pixels in practice) to the wrong side of the infinite plane, the resulting increase of $||\mathbf{e}_j^i||$ may reach hundreds of pixels due to the great distance between antipodal projections (Section 2.3). Although useful for bundle adjustment convergence in our context, such plane crossings are difficult for an LM iteration which should decrease the cost function E (Section 3.1).

We see that infinite plane problems cannot be ignored for error $\mathbf{e}_j^i$ in our context: catadioptric cameras with fields of view greater than 180°, scenes with distant 3D points, approximate calibration knowledge and image noise. In particular, point parametrization (x, y, z) is not sufficient and the use of the standard 2D reprojection error is not straightforward for bundle adjustment. A condition for an ideal error is C^2 continuity with 3D point parametrization (x, y, z, t), even at the infinite plane.

3.3 Image Error

This part presents an image error based on the standard 2D reprojection error of a catadioptric camera. We aim to improve the smoothness of this error for the crossing of the infinite plane (Section 3.2).

Let $\mathbf{X}_j$ be the four homogeneous coordinates of the (finite) j^{th} point in the world coordinates. If the model is central, we define

$$\mathbf{e}_j^{i+} = p_c^i(\mathbf{X}_j) - \mathbf{u}_j^i, \quad \mathbf{e}_j^{i-} = p_c^i(-\mathbf{X}_j) - \mathbf{u}_j^i \tag{8}$$

where $p_c^i(.)$ and $p_c^i(-.)$ are the antipodal projection functions of the i^{th} camera (defined by Equations 1 and 2). If the model is non-central, we define

$$\mathbf{e}_j^{i+} = p_{nc}^{i+}(\mathbf{X}_j) - \mathbf{u}_j^i, \quad \mathbf{e}_j^{i-} = p_{nc}^{i-}(\mathbf{X}_j) - \mathbf{u}_j^i \tag{9}$$

where p_{nc}^{i+} and p_{nc}^{i-} are the antipodal projection functions of the i^{th} camera (defined by Equations 4 and 5).

At first glance, the error function $\mathbf{e}_j^i$ might be defined by the standard 2D reprojection error: $\mathbf{e}_j^i = \mathbf{e}_j^{i+}$. However, this function is discontinuous at the infinite plane (Section 3.2). We propose

$$\mathbf{e}_j^i = \begin{cases} \mathbf{e}_j^{i+}, \text{ if } ||\mathbf{e}_j^{i+}|| < ||\mathbf{e}_j^{i-}|| \text{ or if } \mathbf{e}_j^{i-} \text{ is not defined} \\ \mathbf{e}_j^{i-}, \text{ if } ||\mathbf{e}_j^{i-}|| < ||\mathbf{e}_j^{i+}|| \text{ or if } \mathbf{e}_j^{i+} \text{ is not defined} \end{cases} \tag{10}$$

and see that $||\mathbf{e}_j^i|| = \min(||\mathbf{e}_j^{i+}||, ||\mathbf{e}_j^{i-}||)$ if both $\mathbf{e}_j^{i+}$ and $\mathbf{e}_j^{i-}$ are defined.

First, we show that $||\mathbf{e}_j^{i+}|| = ||\mathbf{e}_j^{i-}||$ is very improbable in practice during the minimization of Equation 6. This would imply values of $||\mathbf{e}_j^i|| = ||\mathbf{e}_j^{i+}|| = ||\mathbf{e}_j^{i-}||$ greater than the radius of the small circle due to the large distance between two antipodal projections (Section 2.3). However, all initial $||\mathbf{e}_j^i||$ are small (less than 2 pixels) and the bundle adjustment attempts to decrease all $||\mathbf{e}_j^i||$ simultaneously by minimizing their sum of squares.

Second, we show that $\mathbf{e}_j^i$ is C^2 continuous at finite point $\mathbf{X}_j$. If $||\mathbf{e}_j^{i+}|| < ||\mathbf{e}_j^{i-}||$ at $\mathbf{X}_j$, this inequality still holds in a neighborhood of $\mathbf{X}_j$ thanks to the continuity of the antipodal functions. Thus, $\mathbf{e}_j^i = \mathbf{e}_j^{i+}$ in this neighborhood thanks to Equation 10. We see that $\mathbf{e}_j^i$ is C^2 continuous at the finite point $\mathbf{X}_j$ just like p_{nc}^{i+} and p_c^i. The proof is similar if $||\mathbf{e}_j^{i-}|| < ||\mathbf{e}_j^{i+}||$ at $\mathbf{X}_j$.

Third, we show that $\mathbf{e}_j^i$ is C^0 continuous at infinite point $\mathbf{X}_\infty$. Let $\mathbf{X}_j$ be a finite point which converges to $\mathbf{X}_\infty$. Section 2.4 shows that the antipodal projections of $\mathbf{X}_j$ converge to the projections $\mathbf{u}_\infty^+$ and $\mathbf{u}_\infty^-$ of $\mathbf{X}_\infty$ in the i^{th} image. If both antipodal projections exist, $\mathbf{e}_j^{i+}$ and $\mathbf{e}_j^{i-}$ converge to $\mathbf{u}_\infty^+ - \mathbf{u}_j^i$ and $\mathbf{u}_\infty^- - \mathbf{u}_j^i$. Thus, $\mathbf{e}_j^i$ converges to the vector among $\mathbf{u}_\infty^+ - \mathbf{u}_j^i$ and $\mathbf{u}_\infty^- - \mathbf{u}_j^i$ which has the smallest modulus (Equation 10): $\mathbf{e}_j^i$ is continuous at $\mathbf{X}_\infty$. If only one antipodal projection exists and converges to $\mathbf{u}_\infty$, the corresponding error $\mathbf{e}_j^{i+}$ (or $\mathbf{e}_j^{i-}$) converges to $\mathbf{u}_\infty - \mathbf{u}_j^i$ and the other error $\mathbf{e}_j^{i-}$ (or $\mathbf{e}_j^{i+}$) is not

defined. Thus, $\mathbf{e}_j^i$ is defined by $\mathbf{e}_j^{i+}$ (or $\mathbf{e}_j^{i-}$) and converges to $\mathbf{u}_\infty - \mathbf{u}_j^i$: $\mathbf{e}_j^i$ is continuous at $\mathbf{X}_\infty$. If no antipodal projection exists, $\mathbf{e}_j^i$ is not defined at $\mathbf{X}_\infty$.

Last, we show that $\mathbf{e}_j^i$ is the standard 2D reprojection error $\mathbf{e}_j^{i+}$ if $\mathbf{X}_j$ is finite and is on the right side of the i^{th} mirror during minimization (i.e. if $\mathbf{X}_j$ has not crossed the infinite plane). In this case, $||\mathbf{e}_j^{i+}||$ is small and the large distance between antipodal projections provides a large $||\mathbf{e}_j^{i-}||$. Thus, $\mathbf{e}_j^i = \mathbf{e}_j^{i+}$.

3.4 Angular Error

We also introduce an angular error for a catadioptric camera: error $\mathbf{e}_j^i$ is such that $||\mathbf{e}_j^i||$ is the angle between two rays. The first one is a back-projected ray obtained by the camera calibration applied to $\mathbf{u}_j^i$. The second one is a ray corresponding to the reprojection by the i^{th} camera of the j^{th} point. Let $\mathbf{d}_j^i$ and $\mathbf{D}_j^i$ be the directions of the first and second ray, respectively.

If the model is central, the two rays start from the camera center $\mathbf{t}^i$ of the i^{th} camera. Let $\mathtt{R}^i$ be its rotation matrix, and $\mathbf{X}_j$ the homogeneous coordinate of the j^{th} point ($\mathbf{t}^i, \mathtt{R}^i, \mathbf{X}_j$ are in the world coordinate system). The definitions of $\mathbf{d}_j^i$ and $\mathbf{D}_j^i$ are straightforward in the camera coordinate system:

$$\mathbf{d}_j^i = \frac{C^{-1}(\mathbf{u}_j^i)}{||C^{-1}(\mathbf{u}_j^i)||}, \quad \mathbf{D}_j^i = \frac{{\mathtt{R}^i}^\top[\mathtt{I}_3| - \mathbf{t}^i]\mathbf{X}_j}{||{\mathtt{R}^i}^\top[\mathtt{I}_3| - \mathbf{t}^i]\mathbf{X}_j||}. \tag{11}$$

If the model is non-central, $\mathbf{s}_j^i, \mathbf{n}_j^i$ and $\mathbf{a}_j^i$ are introduced. Point $\mathbf{s}_j^i$ is the intersection of the mirror surface and the back-projected ray by the perspective camera of $\mathbf{u}_j^i$. We also define $\mathbf{n}_j^i$ as the mirror normal at $\mathbf{s}_j^i$ and $\mathbf{a}_j^i$ as the direction of the back-projected ray (pointing outside the mirror). In this non-central context, we redefine $\mathtt{R}^i$ as the orientation and $\mathbf{t}^i$ as the origin of the i^{th} mirror coordinate system in the world coordinate system. The first ray is a half line starting from $\mathbf{s}_j^i$ with the direction $\mathbf{d}_j^i$ defined by the reflection law. Both $\mathbf{s}_j^i$ and $\mathbf{d}_j^i$ are expressed in the i^{th} mirror coordinates, and are fixed by $\mathbf{u}_j^i$ and all parameters of the perspective camera. The second ray is the half line starting from $\mathbf{s}_j^i$ towards $\mathbf{X}_j$. The expressions of $\mathbf{d}_j^i$ and $\mathbf{D}_j^i$ are easy given $\mathbf{s}_j^i, \mathbf{n}_j^i, \mathbf{a}_j^i$ and X_j^t the fourth homogeneous coordinate of the j^{th} point:

$$\mathbf{d}_j^i = \frac{2({\mathbf{n}_j^i}^\top\mathbf{a}_j^i)\mathbf{n}_j^i - \mathbf{a}_j^i}{||2({\mathbf{n}_j^i}^\top\mathbf{a}_j^i)\mathbf{n}_j^i - \mathbf{a}_j^i||}, \quad \mathbf{D}_j^i = \frac{{\mathtt{R}^i}^\top[\mathtt{I}_3| - \mathbf{t}^i]\mathbf{X}_j - X_j^t\mathbf{s}_j^i}{||{\mathtt{R}^i}^\top[\mathtt{I}_3| - \mathbf{t}^i]\mathbf{X}_j - X_j^t\mathbf{s}_j^i||}. \tag{12}$$

Now, the main subject in this part is the choice of function $\mathbf{e}_j^i$ with the C^2

continuity (Section 3.1) even if $\mathbf{X}_j$ is at the infinite plane (Section 3.2). At first glance, we might choose the 1D error $\mathbf{e}_j^i = \arccos(\mathbf{d}_j^i.\mathbf{D}_j^i)$. Unfortunately, Appendix E shows that this function is not C^1 continuous if $\mathbf{e}_j^i = 0$. This C^1 discontinuity at the exact solution complicates the convergence in practice. A second try might be $\mathbf{e}_j^i = f(\mathbf{d}_j^i.\mathbf{D}_j^i)$ with f a decreasing C^2 continuous function such that $f(1) = 0$, if we accept that $\mathbf{e}_j^i$ is not an angle. The resulting $\mathbf{e}_j^i$ is C^2 continuous and $\mathbf{e}_j^i \geq 0$. We deduce that $\mathbf{e}_j^i$ has a local extrema if $\mathbf{e}_j^i = 0$: the Jacobian of $\mathbf{e}_j^i$ is zero here. The convergence rate of LM is reduced in this context (Section 3.1).

Our final $\mathbf{e}_j^i$ proposition does not have these problems and is defined by

$$\mathbf{e}_j^i = \pi_2(\mathtt{R}_j^i \mathbf{D}_j^i) \text{ with } \mathtt{R}_j^i \text{ a rotation such that } \mathtt{R}_j^i \mathbf{d}_j^i = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{13}$$

Projection π_2 was defined in Equations 3. Since $||\pi_2(x, y, z)|| = \sqrt{\frac{x^2+y^2}{z^2}}$ is the tangent of the angle between $(\,0 \quad 0 \quad 1\,)^\top$ and $(\,x \quad y \quad z\,)^\top$, we see that

$$||\mathbf{e}_j^i||^2 = \tan^2(\alpha_j^i) \text{ with } \alpha_j^i = \mathit{angle}(\mathbf{d}_j^i, \mathbf{D}_j^i).$$

This result is independent of the choice of $\mathtt{R}_j^i$. The resulting cost function E (defined in Equation 6) is a sum of squared tangents of angles between rays. Error $\mathbf{e}_j^i$ is well defined (i.e. $\mathit{angle}(\mathbf{d}_j^i, \mathbf{D}_j^i) \neq \frac{\pi}{2}[\pi]$) and E is a good approximation of the sum of the squared angles in our context since these angles are small in practice. Furthermore,

$$\mathbf{e}_j^i = \pi_2(\mathtt{R}_j^i {\mathtt{R}^i}^\top [\mathtt{I}_3 | - \mathbf{t}^i - \mathtt{R}^i \mathbf{s}_j^i] \mathbf{X}_j) \tag{14}$$

since the projection π_2 cancels the scale of $\mathbf{D}_j^i$ defined in Equations 11 and 12 ($\mathbf{s}_j^i = 0$ in the central case). We recognize a projection by a perspective camera such that the orientation and center are parametrized. So, $\mathbf{e}_j^i$ inherits all high smoothness properties of the perspective camera: $\mathbf{e}_j^i$ is C^2 continuous even if $\mathbf{X}_j$ is in the infinite plane. This smoothness is obvious if the calibration is not refined by the BA. If it is refined by BA, we should assume that the mirror profile is such that $\mathbf{d}_j^i$ (and $\mathbf{s}_j^i$ in the non-central case) are C^2 continuous.

4 Overview of the Automatic Method

Now, the automatic Structure from Motion method is described for both catadioptric camera models using bundle adjustments (BA) proposed in Sections 3.3 and 3.4. First, the geometry of the image sequence is estimated for the central model. Then, a second estimation using the non-central model is obtained. In both cases, an approximate and given calibration is refined with the 3D. Many technical details (useful for possible re-implementers) about matching and initialization steps are also given in the Appendix.

4.1 Assumptions for Calibration Initializations

The following assumption is required: a surface-of-revolution mirror, whose lower and upper circular cross sections are visible, is placed in front of a natural perspective camera (zero skew, aspect ratio set to 1). Furthermore, the perspective camera must point towards the mirror and the perspective center is in the immediate neighborhood of the mirror symmetry axis.

In this context, the projections of the two mirror circular cross sections are approximated by two circles in images. The central model only requires the knowledge of these two circles and the radially symmetric approximation. The non-central model only requires the knowledge of the mirror profile. The assumptions above are needed for the calibration initialization steps in Sections 4.2 and 4.5. These steps are not the main subjects of this paper, and more general methods are available if we assume that the projections of the two mirror circular cross sections are general ellipses (e.g. [20] in the central case and [27] in the non-central case).

4.2 Central Calibration Initialization

First, the large and small circles in each catadioptric image are detected and estimated using RANSAC and Levenberg-Marquardt methods applied on the vertices of regularly polygonized contours. Then, the central calibration C (Equation 2) is initialized as follows. We define $r(\alpha)$ by the linear function such that $r(\alpha_{up}) = r^i_{up}$ and $r(\alpha_{down}) = r^i_{down}$. The image circle radii r^i_{down}, r^i_{up} are obtained in the circle estimation step, and the field of view angles $\alpha_{down}, \alpha_{up}$ are given by the mirror manufacturer. These angles are not exactly known since they depend on the relative position between the mirror and the pinhole camera.

Harris points are detected and matched for each pair of consecutive images in the sequence using correlation without any epipolar constraint. The corresponding ray directions are also obtained from the central calibration initialization. Then, the essential matrices for these pairs are estimated by RANSAC (using the 7 point algorithm [11]) and refined by Levenberg-Marquardt. 3D points are also reconstructed for each pair. Many of these points are tracked in three images, and they are used to initialize the relative 3D scale between two consecutive image pairs. Now, points in 3 views are reconstructed and we obtain the reconstruction of each triple of consecutive images. More details about these matching and estimation steps are given in Appendix A and B.

Last, the full sequence geometry is obtained by many BAs applied in a hierarchical framework to merge all partial geometries [11]. Once the geometries of the two camera sub-sequences $1 \cdots \frac{n}{2}, \frac{n}{2}+1$ and $\frac{n}{2}, \frac{n}{2}+1, \cdots n$ are estimated, the latter is mapped in the coordinate system of the former thanks to the two common cameras $\frac{n}{2}, \frac{n}{2}+1$, and the resulting sequence $1 \cdots n$ is refined by a BA. The angular error (Equations 11 and 13) is preferred for these BAs since we found its convergence more robust in practice. A final BA with image error (Equations 8 and 10) is applied to the full sequence to obtain a MLE with the common Gaussian model in images (Section 1.3).

4.4 Central Reconstruction and Calibration Refinements

Once the full central geometry is obtained for the approximate (linear) function $r(\alpha)$ defined above, $r(\alpha)$ is redefined as a cubic polynomial whose the 4 coefficients should be estimated. We assume that the central calibration is constant and apply an additional BA using image error (Equations 8 and 10) to estimate the $4 + 6c + 3p$ parameters of the sequence (c is the number of cameras, p is the number of 3D points).

4.5 Non-Central Calibration Initialization

The parameters $\mathtt{K}_p, \mathtt{R}_p$ and $\mathbf{t}_p$ of the perspective camera (Section 2.2) are initialized from the detected large and small circles. These circles are the images of the known mirror circular cross sections with radii r_{up} and r_{down} in z-plane $z = z_{up}$ and $z = z_{down}$, respectively. From Section 4.1, we have $\mathtt{R}_p \approx \mathtt{I}_3$, $\mathbf{t}_p = \begin{pmatrix} -x_p & -y_p & -z_p \end{pmatrix}^\top$ where $\max(|x_p|, |y_p|) \ll z_p$, and $\mathtt{K}_p$ is the 3×3 diagonal matrix with focal length f_p.

First, an approximate value of z_p is obtained by measuring the distance between the mirror and the perspective camera. Second, f_p is estimated assuming $x_p = y_p = 0$ by the Thales relation $r^i_{up}/f_p = r_{up}/(z_p + z_{up})$. Third, $\mathtt{R}_p$ and x_p, y_p are estimated by projecting the circular cross section centers. More details are given in Appendix C.

4.6 Non-Central Reconstruction Initialization

The reconstruction with the non-central model is initialized from the reconstruction with the central model. A first possibility is the approximation of each non-central camera by a central camera such that the center is the mirror apex: the i^{th} mirror coordinate system is defined by the pose $(\mathbf{t}^i, \mathtt{R}^i)$ of the i^{th} central camera, and the j^{th} 3D point $\mathbf{X}_j$ is retained as such. The 3D scale factor of the scene, in fact, is not a free parameter as in the central case since the non-central image projections are dependent on it. We choose the initial 3D scale factor by multiplying $\mathbf{t}^i$ and $\mathbf{X}_j$ by $\lambda \in \Re$ such that $d = \frac{1}{c-1}\sum_{i=1}^{c-1} ||\lambda \mathbf{t}^i - \lambda \mathbf{t}^{i-1}||$ is physically plausible. The value of d is obtained by an estimate of the step length between two consecutive images.

Then, a BA is applied to refine the $6c + 3p$ parameters of the non-central reconstruction. We found that the angular error (Equations 12 and 13) is better than the image error (Equations 9 and 10) to start parameter refinement: the convergence is more robust and one LM iteration is about 3 times faster. Last, we finish parameter refinement by applying a BA with the image error to obtain a MLE of the geometry with the common Gaussian model in images (Section 1.3).

4.7 Non-Central Reconstruction and Calibration Refinements

Once the full non-central geometry is obtained from the method above, additional parameters among $\mathtt{K}_p, \mathtt{R}_p$ and $\mathbf{t}_p$ are selected to be refined. We assume that $\mathbf{t}_p$ and f_p are constant during the camera motion in the scene, but the orientation $\mathtt{R}_p$ is perturbed around $\mathtt{I}_3$ (slight rotations of the perspective camera are possible around a screw). Thanks to the mirror symmetry around the z-axis (Section 2.2), $\mathtt{R}_p$ has only two parameters θ_x and θ_y: $\mathtt{R}_p = R_x(\theta_x)R_y(\theta_y)$ with rotations R_x and R_y around x and y-axis of the mirror coordinate system. The number of additional parameters is $k = 1 + 3 + 2c$. Finally, an additional BA with image error (Equations 9 and 10) is applied to refine the $k + 6c + 3p$ parameters of the sequence.

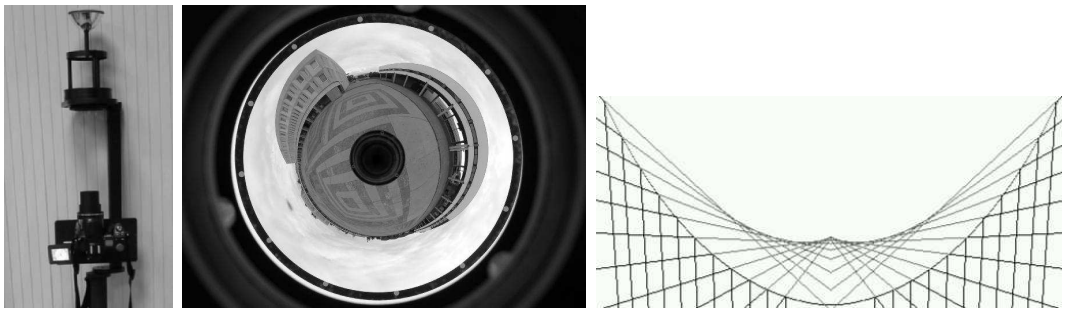

Fig. 5. Left: the 360 One VR (Kaidan) mirror with the Coolpix 8700 (Nikon) camera, mounted on a monopod. Middle: the view field is 360° in the horizontal plane and about 50° above and below if the camera is pointing toward the sky. Right: the profile of the mirror caustic for $z_p = 48$ cm and $x_p = y_p = 0$.

5 Experiments

After the description of the experimental context in Section 5.1, specific results are given for the central model in Section 5.2 and for the non-central model in Section 5.3. Both models are compared with many criteria including accuracy and uncertainty in Section 5.4. Section 5.5 compares reconstructions obtained by minimizing angular and image errors. A piecewise planar 3D model and some technical details are also given in Sections 5.6 and 5.7.

5.1 Experimental Context

The user moves along a trajectory on the ground with the omnidirectional system mounted on a monopod, alternating a step forward and a shot. We use a cheap system designed for panoramic picture generation and image-based rendering from a single view point, given a single shot of the scene (see Figure 5). The mirror is not a quadric and has a known profile (Equation 17) such that we have an approximately uniform resolution in the image radial direction ($r(\alpha)$ of Equation 2 is linear). Such a catadioptric camera/system is called an "equiangular" camera, it is not a central camera, and the size of the caustic profile [9] is about half the size of the mirror profile (right of Figure 5). The symmetry axes of camera and mirror are not exactly the same, since the axes alignment is manually adjusted and visually checked.

5.2 Central Model

Panoramic images obtained from one image of the tested sequences are shown in Figure 6. Top views of resulting reconstructions by methods described in Sections 4.2, and 4.3 are shown in Figure 7 and 11. These results are obtained

Fig. 6. Panoramic images obtained from one omnidirectional image of sequences Fountain, House, and Road. They are given to help understand the scenes, but they are not used by the methods.

Fig. 7. Top views of Fountain (38 views, 5857 points) and House (112 views, 15504 points) central reconstructions. The Road (54 views, 10178 points) reconstruction is shown in Figure 11. Several of these points are difficult to reconstruct accurately if they are distant or roughly aligned with the cameras from which they are reconstructed. These results are similar for central (with linear and cubic $r(\alpha)$ functions) and non-central models.

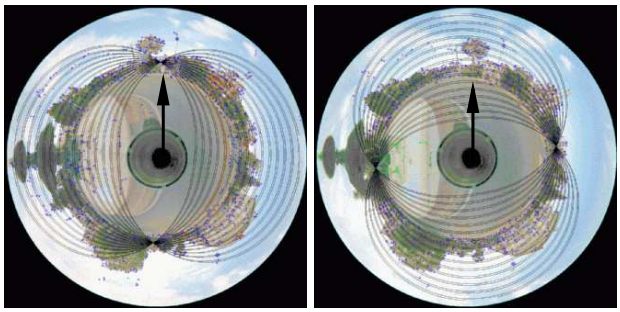

Fig. 8. Epipolar curves for images of two estimated pair-wise central geometries of the Fountain sequence, and true camera motion indicated by the black arrows. The geometry estimation is obviously incorrect on the right since the epipoles do not agree with the true camera motion (they do on the left). Such blunders are corrected by the three view calculations.

with a linear calibration function $r(\alpha)$ defined by the field of view angles $\alpha_{up} = 40°, \alpha_{down} = 140°$ given by the mirror manufacturer.

5.2.1 Fountain

The Fountain sequence is composed of 38 images of a background city and a close-up fountain at the center of a traffic circle. We have found that about 50% of recovered essential matrices are obviously incorrect for this sequence, since the epipoles are roughly orthogonal to the camera motion as shown in Figure 8. However, the 3-view calculations remove all blunders of 2-view calculations thanks to the 3-view selection of 3D points (Appendix B) and angular bundle adjustment with inlier update (Section 5.7). The recovered camera motion is smooth and circular around the fountain. This result is consistent with our knowledge of the true camera motion.

Both unclosed and closed versions of this sequence are reconstructed. The unclosed version is obtained by duplication at the sequence end of the first image, and is useful since it provides information about the pose accuracy by measuring the gap between both sequence ends (this gap is significant of the drift of the unclosed reconstruction process). The distance $||\mathbf{t}^0 - \mathbf{t}^{38}||$ between both sequence ends is 0.01 times the trajectory diameter, and the angle of the relative orientation $\mathtt{R}^{38}(\mathtt{R}^0)^{-1}$ is 0.63°. The closed version is considered in the rest of the paper (including Figures and Tables).

5.2.2 Road and House

The Road sequence is composed of 54 images taken along a little road on flat ground. The background includes buildings, parking and fir trees. All recovered essential matrices seem to be correct according to the positions of the epipoles. The House sequence is composed of 112 images taken in a cosy

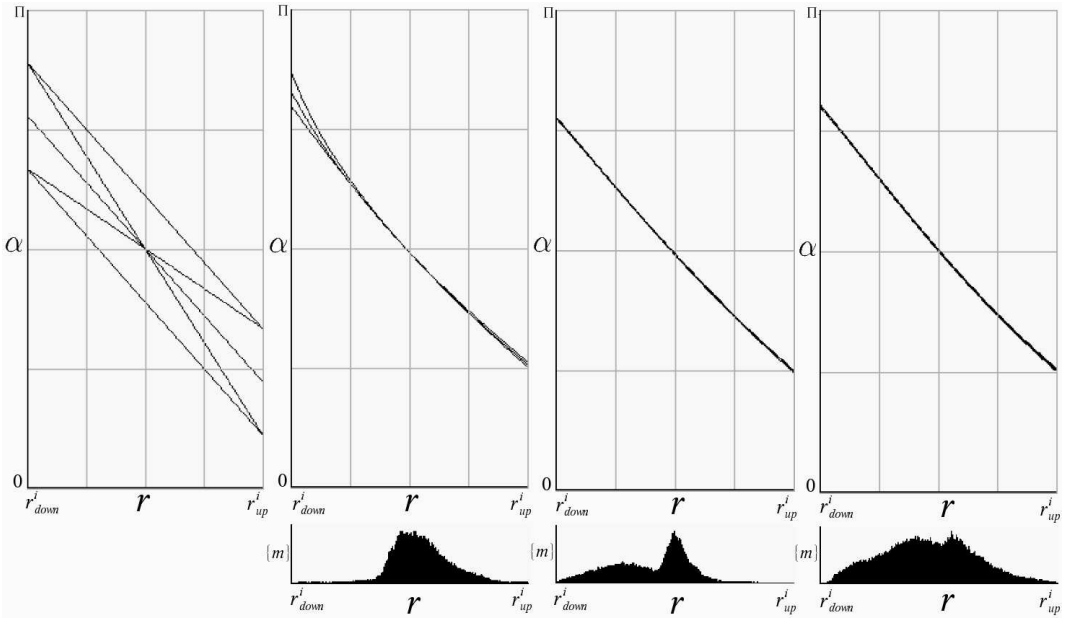

Fig. 9. From left to right: many initial $\alpha(r)$, and the resulting refined $\alpha(r)$ for the Fountain, Road and House sequences, respectively. Function $\alpha(r)$ is the reciprocal function of the radial function $r(\alpha)$ involved in the central calibration (Equation 2). The ranges are always $[r^i_{down}, r^i_{up}] \times [0, \pi]$. The r-distribution of inlier matches $\{m\}$ is also given by histogram for each sequence, at the bottom.

house, starting in the living room, crossing the lobby, having a loop in the kitchen, re-crossing the lobby and entering a bedroom.

5.2.3 Robustness to the given Field of View Angles

This section shows the robustness of the central reconstruction methods to linear calibration function $r(\alpha)$ defined by rough values of α_{up} and α_{down} (Section 4.2). This is useful in practice, since the field of view angles are sometimes unknown or inaccurate (they depend on the relative position between the mirror and the pinhole camera). We also experiment the $r(\alpha)$ refinement.

The experiments are summarized in Figure 9. For each $r(\alpha)$ initialization defined by $(\alpha_{up}, \alpha_{down}) \in \{(40, 140), (20, 160), (60, 120), (20, 120), (60, 120)\}$, the central methods are applied. The reconstruction initialization (Section 4.3) fails twice for the Fountain with these large inaccuracies of $\pm 20°$. No failure occurs for the easier case $\pm 10°$. The last step is the simultaneous reconstruction and calibration refinements (Section 4.4). Function $r(\alpha)$ is redefined as a cubic polynomial and is estimated using bundle adjustment: the four polynomial coefficients are new unknowns, and the set of inliers is updated four times during the minimization. The recovered $r(\alpha)$ are shown on the right of Figure 9 for each initial $(\alpha_{up}, \alpha_{down})$ and each of the 3 sequences. The exact $r(\alpha)$ does not exist since the catadioptric camera is non-central: it depends on the

depth of points. It is also dependent on the settings of the perspective camera which are slightly different for the 3 sequences. However, we assume that the expected results are near the linear $r(\alpha)$ defined by the manufacturer angles. With this in mind, we see that the calibration improvement is significant. Furthermore, the $r(\alpha)$ refinements are stable since the calibration curves are very similar for each sequence, except in the neighborhood of the small circle (radius r^i_{down}) for the Fountain images. This default is probably due to the high lack of image matches that we observed near the small circle in the whole sequence. The means of recovered $(\alpha_{up}, \alpha_{down})$ are $(46°, 149°), (44°, 139°), (45°, 144°)$ for Fountain, Road and House. These angles are slightly greater than the given manufacturer angles $(40°, 140°)$.

5.3 Non-Central Model

Non-central methods are also applied to enforce the mirror knowledge in the geometry estimation (the mirror profile is defined by Equation 17). Calibration and reconstruction are initialized thanks to estimates of z_p and the 3D scale factor of the scene as mentioned in Section 4.5 and Section 4.6, respectively. The largest z_p available with the monopod is chosen to increase the depth of field and obtain the entire mirror in sharp focus with the perspective camera. We measure $z_p = 48$ cm between the mirror and the camera. The approximate trajectory lengths of Fountain, Road, and House sequences are 16, 42 and 22 meters, respectively. The non-central reconstructions are qualitatively similar to the central ones in Figures 7 and 11. The experiments show that the refinements of calibration and 3D scale are difficult in our context. The reasons are given in Section 5.3.1 for the calibration and in Section 5.3.2 for the 3D scale factor.

5.3.1 Calibration Refinements

As mentioned in Section 4.7, there are $1+3+2c$ parameters of the perspective camera to be refined: one focal length f_p, one center $\mathbf{t}_p = (\,-x_p \quad -y_p \quad -z_p\,)^\top$ and many rotations $\mathtt{R}_p = R_x(\theta_x)R_y(\theta_y)$. Obviously, these parameters are added to the $6c + 3p$ parameters of the 3D in the bundle adjustment.

We note that the mirror size (radius 3.7 cm and axis length 3.5 cm) is small in comparison with the distance between the pinhole center and mirror ($z_p = 48$ cm). In this context, small perturbations of θ_x, θ_y and f_p are almost compensated by certain perturbations of $\mathbf{t}_p$ to remain the non-central unchanged projection of a 3D point. More precisely, we have a rotation-translation ambiguity [1]: it is difficult to refine $\mathtt{R}_p$ and the x-y components of $\mathbf{t}_p$ simultaneously. Also, we have ambiguity between zoom and translation

	small scale scene ($r_t = 25$ cm)					large scale scene ($r_t = 2.5$ m)				
s_{init}	0.5	.707	1	1.41	2	0.5	.707	1	1.41	2
$s_{refined}$	.950	.956	.990	.993	.995	.653	.812	.908	1.19	1.69
RMS	.983	.974	.971	.970	.969	.967	.967	.967	.969	.970

Table 1
The ground truth reconstructions are perturbed by initial homothety s_{init} and image Gaussian noise of $\sigma = 1$ pixel. For each s_{init}, a homothety $s_{refined}$ is estimated by non-central bundle adjustment. Homothety $s_{refined}$ is the ratio between the recovered scale factor and its true value (the ideal result is $s_{refined} = 1$). The small scale scene has the best values of $s_{refined}$.

along the focal axis [4]: it is difficult to simultaneously refine f_p and the component of $\mathbf{t}_p$ along the third column of $\mathtt{R}_p$. Obviously, these ambiguities are inherent to the estimation problem (a different method will not remove them).

In the experiments, convergence is slow and the final $f_p, \mathbf{t}_p, \mathtt{R}_p$ values are similar to their initial values. The ambiguities are confirmed by high correlation coefficients for our sequences (obtained from the covariance matrix of the estimated parameters [26]): we always have $|\sigma_{f_p,z_p}| \geq 0.999$, and $|\sigma_{\theta_x,y_p}| \geq 0.8, |\sigma_{\theta_y,x_p}| \geq 0.8$ in the majority of cases.

5.3.2 3D Scale factor Estimation

Refinement of the 3D scale factor of the scene by bundle adjustment is theoretically possible using a non-central catadioptric system, since the non-central image projection changes with the scale factor. However, experiments on our image sequences show that the scales recovered by the non-central reconstruction initialization and refinement (Sections 4.6 and 4.7) are similar to their initial values. The reason is the following: a scale change of the scene (including the mirror trajectory) does not lead to significant changes to the projections of any 3D point distant enough from the mirrors.

The following synthetic experiment shows that it is more difficult to estimate the 3D scale factor for large scale scenes than small scale scenes. Two ground truth reconstructions are defined by half turn of a Fountain-like scene with trajectory radius r_t, mirror orientations $\mathtt{R}^i$ perturbed around $\mathtt{I}_3$, 20 camera poses and 1000 points well distributed in 3D and 2D spaces. The only 3D difference between both reconstructions is the 3D scale of the scene points and the mirror trajectory defined by $r_t = 2.5$ m and $r_t = 25$ cm. The perspective parameters $\mathtt{K}_p, \mathtt{R}_p, \mathbf{t}_p$ are the same and the (exact) image projections are different. First, image projections are corrupted by a Gaussian noise of $\sigma = 1$ pixel, and all mirror locations $\mathbf{t}^i$ and points $\mathbf{X}_j$ are multiplied by an initial factor s_{init}. Second, the bundle adjustments of Section 4.6 are applied

Calib.	initial central	initial non-central	refined central	refined non-central
Criteria	#3D,#2D,RMS	#3D,#2D,RMS	#3D,#2D,RMS	#3D,#2D,RMS
Fountain	5857,31028,0.84	6221,33262,0.78	5953,32214,0.77	6223,33876,0.76
Road	10178,50225,0.87	11312,56862,0.82	10814,55297,0.77	11313,57148,0.79
House	15504,75100,0.83	16432,80169,0.78	15800,77883,0.76	16447,80311,0.76
d.o.f.	$6c + 3p - 7$	$6c + 3p - 6$	$4 + 6c + 3p - 7$	$4 + 8c + 3p - 6$

Table 2
The RMS in pixels, the numbers of 3D reconstructed and 2D detected points (inliers) for four reconstructions: central reconstructions with initial and refined calibrations, non-central reconstructions with initial and refined calibrations. The degree of freedom (d.o.f.) depends on the numbers c and p of cameras and 3D points.

to these perturbed reconstructions enforcing the exact perspective parameters ($z_p = 48$ cm) and taking into account the outliers. Table 1 shows for many values of s_{init} the resulting RMS (pixels) and ratio $s_{refined}$ between the recovered scale factor and its exact value. We note that the RMS has no clear minimum for the large scale scene with $r_t = 2.5$ m, and that the scale factor estimations are best for the small scale scene with $r_t = 25$ cm.

In practice, we have abandoned attemps to estimate the 3D scale factor accurately: the measure of distance between two scene points or camera centers is too difficult with this catadioptric system and usual scenes.

5.4 Performance Comparisons Between Central and Non-Central Models

Quantitative comparisons between real 3D reconstructions using the central and non-central models are given.

5.4.1 Consistency

Table 2 shows consistencies between many reconstructions and the multi-view matching for the sequences Fountain, Road and House. There are four reconstructions: central reconstructions with initial (Section 4.3) and refined (Section 4.4) calibrations, non-central reconstructions with initial (Section 4.6) and refined (Section 4.7) calibrations. The consistency criteria are the RMS, the numbers of 3D and 2D points which are consistent (3D reconstructed and 2D detected points $\mathbf{X}_j$ and $\mathbf{u}_j^i$ are consistent if the standard 2D reprojection error $||\mathbf{e}_j^i||$ is less than 2 pixels). The non-central reconstructions have the best consistencies: improvements about 5% (sometimes 12% for the Road) are obtained for the numbers of 3D and 2D points, with slightly slower RMS. We also see that the consistency improvements by calibration refinement are non

Fig. 10. A panoramic image from the "controlled" sequence.

negligible for the central model. They are negligible for the non-central model.

5.4.2 Difference

Two reconstructions "a" and "b" are compared as follows. The camera location difference and 3D point difference are respectively

$$E_t^{(a,b)} = \sqrt{\frac{1}{I}\sum_i ||S(\mathbf{t}_b^i) - \mathbf{t}_a^i||^2}, \quad E_x^{(a,b)} = \sqrt{\frac{1}{J}\sum_j \frac{||S(\mathbf{X}_j^b) - \mathbf{X}_j^a||^2}{||\mathbf{t}_a^{i(j)} - \mathbf{X}_j^a||^2}} \tag{15}$$

with S the similarity transformation minimizing $E_t^{(a,b)}$, I the number of cameras, J the number of 3D points, and $i(j)$ the index of the closest $\mathbf{t}_a^i$ to the j^{th} point $\mathbf{X}_j^a$. $\mathbf{t}_a^i$ is the apex of the i^{th} mirror in the non-central case and the i^{th} camera center in the central case.

The 3D differences between central $(\mathbf{t}_c^i, \mathbf{X}_j^c)$ and non-central $(\mathbf{t}_{nc}^i, \mathbf{X}_j^{nc})$ reconstructions are the followings. We obtain $E_t^{(nc,c)} = 0.79$ cm, $E_x^{(nc,c)} = 0.027$ for the Fountain (respectively, $E_t^{(nc,c)} = 2.6$ cm, $E_x^{(nc,c)} = 0.017$ and $E_t^{(nc,c)} = 2.75$ cm, $E_x^{(nc,c)} = 0.054$ for the Road and House).

5.4.3 Pose accuracy

A real sequence (Figure 10) is taken in an indoor controlled environment: the motion of the catadioptric system is measured on a rail, in a $7m \times 5m \times 3m$ room. The trajectory is a 1 meter long straight line by translation, with 6 equidistant and aligned poses. The location errors are $E_t^{(g,c)}$ and $E_t^{(g,nc)}$ with $\mathbf{t}_g^i$ the location ground truth and $E_t^{(.,.)}$ defined in Equations 15. We obtain $E_t^{(g,c)} = 1.1$ mm and $E_t^{(g,nc)} = 1.2$ mm for the central and non-central models, respectively. Both models provides similar and good accuracies for the location estimation in this context. The angle $E_r = \max_{i,j}\{\arccos(\frac{1}{2}(trace(\mathtt{R}^i(\mathtt{R}^j)^{-1}) - 1))\}$ is our orientation error since the ground truth of $\mathtt{R}^i$ is unknown and maintained constant. The results are $E_r^c = 0.5°$ and $E_r^{nc} = 0.35°$.

	Length	Gauge constraints	u_c	$u_p^{0/4}$	$u_p^{1/4}$	$u_p^{2/4}$	$u_p^{3/4}$	$u_p^{4/4}$
Fountain	16 m	fixed $\mathtt{R}^0, \mathbf{t}^0, \mathbf{t}_x^{14}$	1.02	22	50.7	102	230	27e+4
Road	42 m	fixed $\mathtt{R}^0, \mathbf{t}^0, \mathbf{t}_x^{53}$	4.60	4.37	7.00	23.1	130	44e+4
House	22 m	fixed $\mathtt{R}^0, \mathbf{t}^0, \mathbf{t}_x^{111}$	7.62	5.17	7.27	12.2	26.7	11e+3

Table 3
Right: the camera center and 3D point uncertainties (cm) for central reconstructions. Left: the lengths (m) of the trajectories and the gauge constraints. The uncertainties are the length of the major semi-axis of the uncertainty ellipsoids for the probability 90%. $u_p^{0/4}, u_p^{1/4}, u_p^{2/4}, u_p^{3/4}, u_p^{4/4}$ are respectively the rank 0/4 (smallest), rank 1/4, rank 2/4 (median), rank 3/4 and rank 4/4 (largest) semi-axis lengths for the 3D points. u_c is the largest semi-axis length for the cameras.

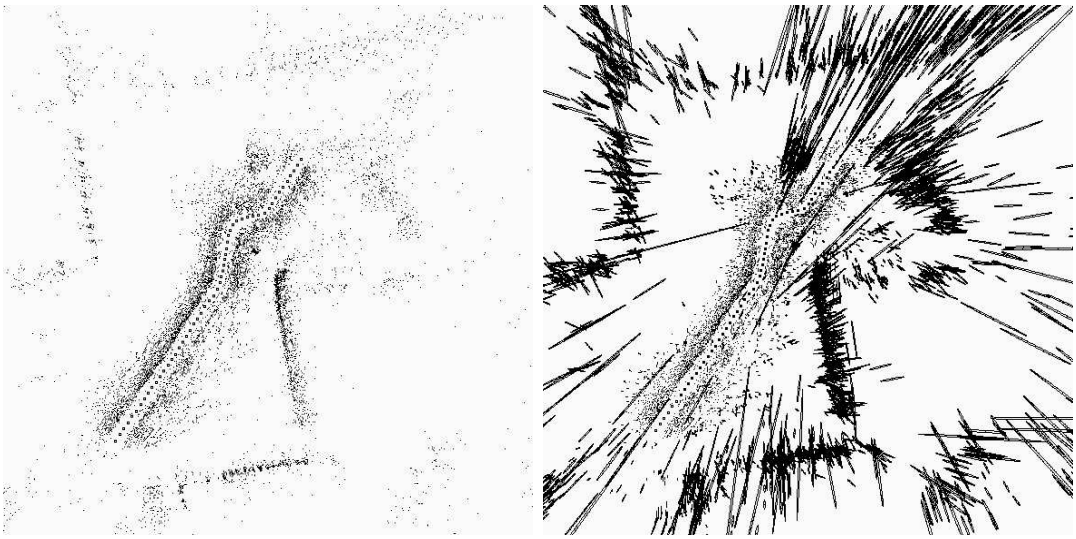

Fig. 11. Top view of the Road central reconstruction (54 views, 10178 points) and the uncertainty ellipsoids of Table 3. Ellipsoids are very long for 3D points which are distant or roughly aligned with the cameras from which they are reconstructed.

5.4.4 Uncertainty

When ground truth is not available, informations about the reconstruction quality are provided by the uncertainty (covariance matrix) estimation of the geometry parameters [11,26]. This estimation requires the common Gaussian model (Section 1.3) for the image error minimized by bundle adjustment.

A trivial camera-based gauge constraints is chosen to obtain minimal parameterizations [26]. Since the central reconstruction is defined up to a similarity transformation (7 d.o.f.), we choose the constraints $\mathtt{R}^0 = \mathtt{I}_3, \mathbf{t}^0 = 0$ and one fixed coordinate of $\mathbf{t}^{i_0}$. The resulting uncertainties for 3D points and camera centers are given in Table 3 (also refer to Figure 11 for the Road sequence). In the non-central case, only $\mathtt{R}^0 = \mathtt{I}_3, \mathbf{t}^0 = 0$ should be sufficient since the global scale of the scene is fixed by the mirror size. In practice, we found that the

medians of the resulting non-central uncertainties are 3-26 greater than the medians of the central uncertainties. The reason is given in Section 5.3.2: the 3D scale factor of the scene cannot be estimated accurately.

5.5 Angular vs. Image Error

The minimization of angle errors (Section 3.4) may appear to be a heuristic choice. It is shown in this part that the results are not so different to those obtained with the more standard minimization of image errors (Section 3.3). We use the ground truth reconstructions introduced in Section 5.3.2 and corrupt the image projections by a Gaussian noise of $\sigma = 1$ pixel. Then, we compare the reconstructions $(\mathbf{t}_A^i, \mathbf{X}_j^A)$ and $(\mathbf{t}_I^i, \mathbf{X}_j^I)$ obtained by a BA minimizing angular and image errors, respectively with the ground truth $(\mathbf{t}_g^i, \mathbf{X}_j^g)$. The BAs are initialized by the noisy ground truth reconstruction, the calibration and inliers are maintained to be the same, and the reconstruction comparisons are done with the measures defined by Equations 15.

The results are $E_t^{(g,A)} = 0.29$ cm, $E_x^{(g,A)} = 0.027$, $E_t^{(g,I)} = 0.28$ cm and $E_x^{(g,I)} = 0.027$ for the non-central and large scale reconstruction of Section 5.3.2. They are similar with the central model and/or the small scale reconstruction.

5.6 Piecewise Planar 3D Modeling

For each planar piece of the targeted 3D model, we choose a catadioptric image of the sequence and define the contour of the piece manually. Then, we select reconstructed points by their projections in the delimited region and estimate the support plane from these points by minimizing a sum of (squared) point-to-plane distances. The Mahalanobis point-to-plane distance [24] is preferred to the Euclidean distance to favor the points with the least uncertainties. In this context, independent point covariances are useful in order to obtain a Maximum Likelihood Estimation of the plane. So these covariances are estimated by inverting each Hessian of the independent ray intersection problems.

A piecewise planar model of the House is shown in Figure 12. It is not difficult to guess where the living-room, the kitchen, the lobby, and the bedroom are. 36 planes are estimated from the non-central reconstruction. Plane accuracy depends on the number of 3D points selected (between 7 and 178), their accuracies, uncertainties, and distributions in images.

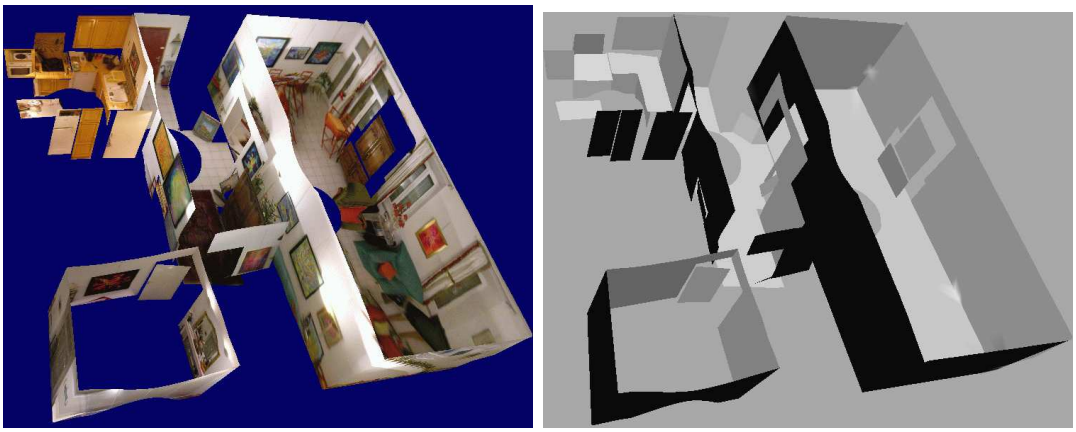

Fig. 12. A piecewise planar model of the House.

5.7 Some Technical Details

Computation times for 1632×1224-images with a P4 2.8GHz/800MHz are about 10 seconds for each single image calculation (points and edge detections, circles), 20 seconds for each image pair calculation (matching by quasi-dense propagation [14], essential matrix estimation) for each sequence, and 5 minutes for the hierarchical reconstruction step applied to the House sequence (total time: 1 hour). About 700-1100 point matches satisfy the epipolar constraint between two consecutive images.

An image point is considered as an outlier for all angular (respectively, image) bundle adjustments if the corresponding error is greater than 0.04 radians (respectively, 2 pixels). The sets of inliers and outliers are updated one time during each bundle adjustment. Accuracy increases when new inliers are discovered and involved in the score to minimize. Furthermore, inliers which becomes outliers after the first round of bundle adjustment are ignored in the second round to gain robustness. Usual and final RMS are about 0.005-0.006 radians (respectively, 0.7-0.8 pixels) with RMS$= \sqrt{E/(\#2D)}$. Function E is defined by Equation 6 and $\#2D$ is the number of $\mathbf{e}_j^i$ involved in E.

All bundle adjustments are implemented using fully analytical derivatives, except the image error in the non-central case which requires an iterative method for M^+ and M^- calculations (more details in Appendix D). One bundle adjustment iteration for the final refinement of the House takes about 1.6 seconds using the non-central image error (7 seconds are also required for M^+ and M^- initialization before all optimizations), and about 0.6 seconds for others bundle adjustments.

6 Conclusion

The methods described in this paper provide an automatic, robust and optimal estimation of the scene structure and camera motion for image sequences acquired by a catadioptric camera. First, we propose bundle adjustments minimizing angle and image errors, by taking care of the targeted smoothness conditions for good convergence. The image error provides a Maximum Likelihood Estimate of the sequence geometry for the common Gaussian model in images, and its smoothness is improved in the infinite plane. The angle error has the ideal smoothness (C^2 continuity, even at the infinite plane). The second contribution is an extensive experimental study in a context that we expect representative and useful for applications: a hand-held and equiangular catadioptric camera moving on the ground. Many experiments are presented about robustness (including the initialization robustness to the given field of view angles), accuracy and uncertainty estimations, performance comparisons between central and non-central models, and piecewise planar 3D modeling from the reconstructed points. The central model provides good approximation of the (real) non-central model. On the other hand, the 3D scale factor is difficult to estimate with the non-central model. The only demonstrated improvement by the non-central model is obtained for the consistency between matching and geometry. We also discuss calibration refinements for both central and non-central models. Last, the 3D reconstruction system is described as a whole.

Future works include applications (image-based modeling and rendering, vehicle localization ...), matching improvements, and camera calibration by enforcing constraints on the 3D scene (especially for the non-central case which has rarely been discussed until now).

Appendix A. Two-View Matching

The usual matching procedure of interest points using correlation cannot be directly applied to the omnidirectional images for two reasons: (1) the matching ambiguity due to the repetitive textures in one image and (2) the geometric distortions between matched patches of two images taken from different view points. Previously published matching methods deal only with one of these problems (e.g. relaxation [28] is used for repetitive textures, regions [18] or appropriate correlation windows [25] for geometric distortions in omnidirectional images).

In the context of the usual camera motions (roughly, translation motions with the pinhole camera pointing toward the sky), we observe that a high pro-

portion of the distortions is compensated for by image rotation around the circle center. The Harris point detector [10] is used because it is invariant to such rotations and it has good detection stability. We also compensate for the rotation in the neighborhood of the detected points before comparing the luminance neighborhood of two points using the ZNCC score (Zero Mean Normalized Cross Correlation). To avoid incorrect matching due to repetitive textures, the following procedure is applied. First, the points of interest of an image are matched to other points of interest in the same image. The result is a "reduced" list of points which are *not* similar to others according to ZNCC in their corresponding search area. Second, the points of the reduced lists of two different images are matched applying the same correlation score, search areas and thresholds. Now the matching errors due to repetitive textures have been greatly reduced, but the current list of matches is very incomplete. Third, this list has been completed thanks to a quasi-dense match propagation [14]: the majority of image pixels are progressively matched using a 2D-disparity gradient limit and the uniqueness constraint, and two interest points roughly satisfying the resulting correspondence mapping between both images are added to the list. The window sizes and lower bound thresholds for ZNCC are the same as in [14]. The resulting list of matched interest points is used in the epipolar geometry estimation step (described in Appendix B).

Appendix B. Two-View and Three-View Central Initialization

More details are given on the geometry initialization described in Section 4.3. Let $\mathbf{d}_j^0$ and $\mathbf{d}_j^1$ be the ray directions (normalized vectors) obtained by the central calibration applied to the j^{th} pair of matched points in images 0 and 1. First, a fundamental matrix $\mathtt{F}$ is obtained with the 7-point algorithm [11] : (1) each pair $(\mathbf{d}_j^0, \mathbf{d}_j^1)$ provides a normalized linear equation for the 9 parameters of $\mathtt{F}$, (2) $\mathtt{F} = \mathtt{F}_1 + \lambda \mathtt{F}_2$ where $\mathtt{F}_1, \mathtt{F}_2$ are two solutions of a 7×9 linear system and (3) $\mathtt{F}$ is obtained by solving a cubic polynomial equation in λ to enforce the constraint $\det(\mathtt{F}) = 0$. Second, an essential matrix $\mathtt{E}$ is obtained from $\mathtt{F}$ by forcing the two largest singular values of $\mathtt{F}$ to be the same by SVD [11]. Third, we count the number of pairs $(\mathbf{d}_j^0, \mathbf{d}_j^1)$ such that the angle between $\mathbf{d}_j^1$ and the normal $\frac{\mathtt{E}\mathbf{d}_j^0}{||\mathtt{E}\mathbf{d}_j^0||}$ of the epipolar plane of $\mathbf{d}_j^0$ is equal to $\frac{\pi}{2}$ up to a threshold t_0. This process is repeated many times by the RANSAC method, and the $\mathtt{E}$ with the largest number of pairs is retained. The $\mathtt{E}$ refinement by Levenberg-Marquardt is straightforward: we use the parametrization $E(\mathbf{t}, \mathtt{R}) = [\mathbf{t}]_\times \mathtt{R}$ and minimize a Longuet-Higgins criterion [5] defined by $LH(\mathbf{t}, \mathtt{R}) = \sum_j (\mathbf{d}_j^1 . \frac{E(\mathbf{t},\mathtt{R})\mathbf{d}_j^0}{||E(\mathbf{t},\mathtt{R})\mathbf{d}_j^0||})^2$. The last step of the two-view geometry initialization is the reconstruction of each point $\mathbf{X}_j$ by minimizing the function $\mathbf{X}_j \rightarrow ||\mathbf{e}_j^0||^2 + ||\mathbf{e}_j^1||^2$ with $\mathbf{e}_j^i$ defined by Equations 11 and 13.

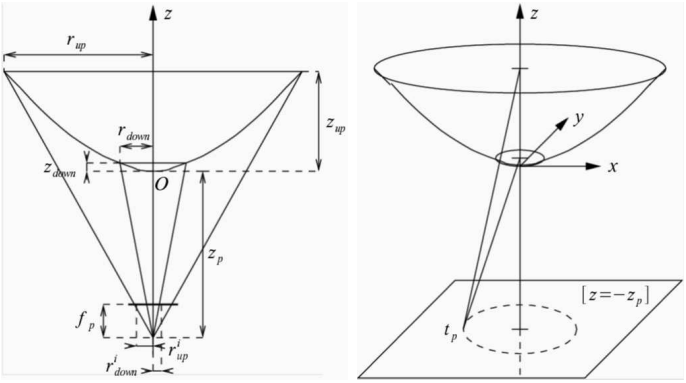

Fig. 13. Left: useful notations for non-central calibration initialization assuming $x_p = y_p = 0$. Right: t_p is located in a circle in the plane $[z = -z_p]$ given z_p and the angle between directions from t_p toward the centers of large and small border circles.

Once the geometries of image pairs $(0, 1)$ and $(1, 2)$ are estimated with the method above, we know the central poses $(\mathbf{t}^0, \mathtt{R}^0), (\mathbf{t}^1, \mathtt{R}^1)$ and $(\lambda\mathbf{t}^2, \mathtt{R}^2)$ of the geometry of the image triple $(0, 1, 2)$ up to the relative 3D scale $\lambda \in \Re$ with $\mathbf{t}^1 = 0$. The 1 point RANSAC below is used to estimate λ. For each $\mathbf{X}_j$ detected in image triple $(0, 1, 2)$ and reconstructed from image pair $(0, 1)$, λ_j is estimated by minimizing the angular error in image 2 defined by $E_j^2(\lambda) = ||\pi_2(\mathtt{R}_j^2{\mathtt{R}^2}^\top[\mathtt{I}_3| - \lambda\mathbf{t}^2]\mathbf{X}_j)||$. This error is derived from Equation 14 with $\mathbf{s}_j^i = 0$. Then, we choose the λ_j with the greatest number of points $\mathbf{X}_{j'}$ such that $E_{j'}^2(\lambda_j)$ is less than a threshold t_1.

The point $\mathbf{X}_j$ is retained in the geometry of the image triple $(0, 1, 2)$ if it is detected in images $0, 1$ and 2 (2 views are not enough for robustness). Furthermore, $\mathbf{X}_j$ is reconstructed by minimizing the angular cost function $\mathbf{X}_j \rightarrow ||\mathbf{e}_j^0||^2 + ||\mathbf{e}_j^1||^2 + ||\mathbf{e}_j^2||^2$ and it should satisfy $\max(||\mathbf{e}_j^0||, ||\mathbf{e}_j^1||, ||\mathbf{e}_j^2||) \leq t_2$ with t_2 a threshold. Last, a bundle adjustment with angular error (Equations 11 and 13) is used to refine the complete geometry of the image triple. Only one angular threshold $t_0 = t_1 = t_2 = 0.04$ (radians) is used in practice.

Appendix C. Non-Central Calibration Initialization

More details on the calibration initialization described in Section 4.5 are given in this Appendix. We estimate $\mathtt{K}_p = diag(f_p, f_p, 1)$, $\mathbf{t}_p = (\,-x_p \quad -y_p \quad -z_p\,)^\top$ and $\mathtt{R}_p$ in the mirror coordinate system. First, an approximate value of z_p is obtained by measuring the distance between the mirror apex and the perspective camera. Second, f_p is estimated assuming $x_p = y_p = 0$ by the Thales relation $r_{up}^i/f_p = r_{up}/(z_p + z_{up})$ (left of Figure 13).

Let $\mathbf{c}_{up}^i$ and $\mathbf{c}_{down}^i$ be the projections by the perspective camera of the centers of

the two mirror circular cross sections. Thanks to the hypotheses (Section 4.1), $\mathbf{c}^i_{up}$ and $\mathbf{c}^i_{down}$ are approximated by the known centers of the detected large and small circles. Since the angle between directions pointing toward the centers of mirror border circles is that between vectors $\mathtt{K}_p^{-1}\mathbf{c}^i_{up}$ and $\mathtt{K}_p^{-1}\mathbf{c}^i_{down}$, we know the radius r_t of the circle in the plane $[z = -z_p]$ where $\mathbf{t}_p$ is located (right of Figure 13). Any $\mathbf{t}_p$ in this circle is possible by the over-parametrization $(\mathtt{R}_p, \mathbf{t}_p, \mathtt{R})$ of the non-central model (Section 2.2).

Now, $\mathtt{R}_p$ and x_p, y_p are estimated by projecting the circular cross section centers:

$$\lambda_u \begin{pmatrix} x_p \\ y_p \\ z_p + z_{up} \end{pmatrix} = \mathtt{R}_p \begin{pmatrix} \mathbf{c}^i_{up} \\ f_p \end{pmatrix}, \lambda_d \begin{pmatrix} x_p \\ y_p \\ z_p + z_{down} \end{pmatrix} = \mathtt{R}_p \begin{pmatrix} \mathbf{c}^i_{down} \\ f_p \end{pmatrix}. \tag{16}$$

These equations and the hypothesis $\mathtt{R}_p \approx \mathtt{I}_3$ imply

$$\lambda_u \approx \frac{f_p}{z_{up} + z_p} < \frac{f_p}{z_{down} + z_p} \approx \lambda_d, \quad \mathbf{c}^i_{up} - \mathbf{c}^i_{down} \approx (\lambda_u - \lambda_d) \begin{pmatrix} x_p \\ y_p \end{pmatrix}.$$

Since any $(\, x_p \quad y_p \,)^\top$ is possible in the circle $x_p^2 + y_p^2 = r_t^2$, we can choose $(\, x_p \quad y_p \,)^\top = \frac{r_t}{||\mathbf{c}^i_{up} - \mathbf{c}^i_{down}||}(\mathbf{c}^i_{down} - \mathbf{c}^i_{up})$ thanks to the hypothesis $\mathtt{R}_p \approx \mathtt{I}_3$. Last, $\mathtt{R}_p$ is estimated from Equations 16

Appendix D. Estimation and Derivation of $M^+(\mathbf{A}, \mathbf{B})$

The bundle adjustment with the non-central image error (Equations 9 and 10) requires efficient estimation and differentiation of $M^+(\mathbf{A}, \mathbf{B})$. This function gives the reflection point on the mirror surface for the ray which goes across points $\mathbf{A}$ and $\mathbf{B}$. Once these computations are done at $(\mathbf{t}_p, \mathbf{X})$ with pinhole center $\mathbf{t}_p$ and scene point $\mathbf{X}$ (in the mirror coordinate system), all derivatives of the projection $p^+_{nc}(\mathbf{X})$ (Equation 4) are analytical according to the Chain Rule. Calculations are very similar for $M^-(\mathbf{A}, \mathbf{B})$ and $p^-_{nc}(\mathbf{X})$ (Equation 5).

The mirror surface is defined by the cylindric parameterization

$$f(r, \theta) = (\, r\cos(\theta) \quad r\sin(\theta) \quad z(r) \,)^\top$$

with the mirror profile

$$z(r) = 0.0287r + 0.218r^2 - 0.0156r^3 + 0.00537r^4, \quad 0 \leq r \leq 3.7 \text{ cm}. \tag{17}$$

Given $\mathbf{A}, \mathbf{B} \in \Re^3$, (r, θ) is estimated such that $f(r, \theta)$ is the reflection point $M^+(\mathbf{A}, \mathbf{B})$ onto the mirror surface. Thus we have

$$n(r,\theta) \wedge \left(\frac{f(r,\theta) - \mathbf{A}}{||f(r,\theta) - \mathbf{A}||} + \frac{f(r,\theta) - \mathbf{B}}{||f(r,\theta) - \mathbf{B}||}\right) = 0$$

by the reflection law, with $n(r, \theta)$ the mirror normal at point $f(r, \theta)$. Only 2 among these 3 equations are independent: (r, θ) is estimated such that $g(r, \theta, \mathbf{A}, \mathbf{B}) = 0$ where

$$g(r,\theta,\mathbf{A},\mathbf{B}) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & z'(r) \end{pmatrix} R_z(-\theta)\left(\frac{f(r,\theta) - \mathbf{A}}{||f(r,\theta) - \mathbf{A}||} + \frac{f(r,\theta) - \mathbf{B}}{||f(r,\theta) - \mathbf{B}||}\right)$$

and $R_z(-\theta)$ is the rotation of angle $-\theta$ around the mirror z-axis.

We obtain (r, θ) using the Gauss-Newton method by minimizing $||g||^2$. The Implicit Function Theorem is also applied to g and provides locally a C^1-function which maps $(\mathbf{A}, \mathbf{B})$ to (r, θ) and its derivation

$$\begin{pmatrix} d_{\mathbf{A}} r & d_{\mathbf{B}} r \\ d_{\mathbf{A}} \theta & d_{\mathbf{B}} \theta \end{pmatrix} = -(d_{(r,\theta)} g)^{-1} d_{(\mathbf{A},\mathbf{B})} g$$

using the notation $d_{\mathbf{X}}\mathbf{Y}$ for the Jacobian of function which maps $\mathbf{X}$ to $\mathbf{Y}$. The value and derivatives of $M^+(\mathbf{A}, \mathbf{B}) = f(r, \theta)$ are deduced by function composition and Chain Rule.

We note that all derivative calculations for one image error (Equation 4) are essentially analytical, since the "numerical part" is greatly reduced to a single 2D Gauss-Newton method.

Appendix E. Against the Exact Angular Error

This appendix shows that the exact angular error

$$e_j^i = \arccos(\mathbf{d}_j^i . \mathbf{D}_j^i), \quad \mathbf{D}_j^i = \frac{\mathtt{R}^{i\top}[\mathtt{I}_3 | -\mathbf{t}^i]\mathbf{X}_j - \mathbf{s}_j^i X_j^t}{||\mathtt{R}^{i\top}[\mathtt{I}_3 | -\mathbf{t}^i]\mathbf{X}_j - \mathbf{s}_j^i X_j^t||}$$

is not C^1 continuous when $e_j^i = 0$. Without loss of generality, the space coordinate system is changed such that $\mathbf{d}_j^i = (\,0 \quad 0 \quad 1\,)^\top$, and we write

$$\mathbf{D}_j^i = \frac{1}{\sqrt{x^2(\alpha) + y^2(\alpha) + z^2(\alpha)}} \begin{pmatrix} x(\alpha) \\ y(\alpha) \\ z(\alpha) \end{pmatrix}$$

with $x(\alpha), y(\alpha), z(\alpha)$ three real C^1 continuous functions with parameter α such that $(\, x(0) \quad y(0) \quad z(0)\,) = (\, 0 \quad 0 \quad 1\,)$. Now, we show that the limit of $\frac{\partial e^i_j}{\partial \alpha}$ is not well defined if α converges to 0.

The Chain Rule provides

$$\frac{\partial e^i_j}{\partial \alpha} = \arccos'(\mathbf{d}^i_j.\mathbf{D}^i_j)\frac{\partial}{\partial \alpha}(\mathbf{d}^i_j.\mathbf{D}^i_j) \text{ with } \arccos'(u) = \frac{-1}{\sqrt{1-u^2}} \text{ if } |u| < 1.$$

Using shortened notations x, y, z for $x(\alpha), y(\alpha), z(\alpha)$, we have

$$\frac{\partial e^i_j}{\partial \alpha} = \frac{-1}{x^2+y^2+z^2}(\frac{\partial z}{\partial \alpha}\sqrt{x^2+y^2} - \frac{z}{\sqrt{x^2+y^2}}(x\frac{\partial x}{\partial \alpha} + y\frac{\partial y}{\partial \alpha})).$$

Since $(\, x(\alpha) \quad y(\alpha) \quad z(\alpha)\,) \approx (\, \frac{\partial x}{\partial \alpha}(0)\alpha \quad \frac{\partial y}{\partial \alpha}(0)\alpha \quad 1\,)$, we obtain

$$\frac{\partial e^i_j}{\partial \alpha} \approx \frac{\alpha}{|\alpha|}\sqrt{(\frac{\partial x}{\partial \alpha})^2(0) + (\frac{\partial y}{\partial \alpha})^2(0)}.$$

Two $\frac{\partial e^i_j}{\partial \alpha}$ limits are obtained: one for each possible α sign.

References

[1] G. Adiv, "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," *TPAMI*, vol. 11, pages 477-489, 1989.

[2] D.G. Aliaga, "Accurate Catadioptric Calibration for Real-time pose estimation in Room-size Environments," *ICCV'01*.

[3] "Boujou," *2d3 Ltd*, http://www.2d3.com, 2000.

[4] S. Bougnoux, "From Projective to Euclidean Space under any practical situation, a criticism of self-calibration," *ICCV'98*.

[5] O.D. Faugeras, "Three-Dimensional Computer Vision - A Geometric Viewpoint," *MIT Press*, 1993.

[6] O.D. Faugeras and Q.T. Luong, "The Geometry of Multiple Images," *MIT Press*, 2001.

[7] C. Geyer and K. Daniilidis, "A unifying Theory for Central Panoramic Systems and Practical Implications," *ECCV'00*.

[8] C. Geyer and K. Daniilidis, "Structure and Motion from Uncalibrated Catadioptric Views," *CVPR'01*.

[9] M. Grossberg and S.K. Nayar, "A general imaging model and a method for finding its parameters," *ICCV'01*.

[10] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Alvey Vision Conf.*, pp. 147-151, 1988.

[11] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," *Cambridge University Press*, 2000.

[12] S.B. Kang and R. Szeliski, "3D Scene Data Recovery Using Omnidirectional Multibaseline Stereo," *IJCV*, vol. 25, no. 2, pp. 167-183, 1997.

[13] S.B. Kang, "Catadioptric Self-Calibration," *CVPR'00*.

[14] M. Lhuillier and L. Quan, "Match Propagation for Image-Based Modeling and Rendering," *TPAMI*, vol. 24, no. 8, pp. 1140-1146, 2002.

[15] M. Lhuillier and L. Quan, "A Quasi-Dense Approach to Surface Reconstruction from Uncalibrated Images" *TPAMI*, vol. 27, no. 3, pp. 418-433, 2005.

[16] M. Lhuillier, "Automatic Structure and Motion using a Catadioptric Camera" *OMNIVIS'05* (workshop).

[17] C.P. Lu, G.D. Hager and E. Mjolsness, "Fast and Globally Convergent Pose Estimation from Video Images," *TPAMI*, vol. 22, no. 6, pp. 610-622, 2000.

[18] J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *BMVC'02*.

[19] K. Madsen, H. Nielsen and O. Tingleff, "Methods for Non-Linear Least Squares Problems Problems," Technical University of Denmark, 2004. Lecture notes.

[20] B. Micusik and T. Pajdla, "Structure from Motion with Wide Circular Field of View Cameras," *TPAMI*, vol. 28, no. 7, pp. 1135-1149, 2006.

[21] D. Nister, O. Naroditsky and J. Bergen, "Visual Odometry," *CVPR'04*.

[22] M. Pollefeys, R. Koch and L. Van Gool, "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters," *ICCV'98*.

[23] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, "Numerical Recipes in C," *Cambridge University Press*, 1988.

[24] K. Schindler and H. Bischof, "On Robust Regression in Photogrammetric Point Clouds," *DAGM'03* (also *LNCS 2781*, pp. 172-178, 2003).

[25] T. Svoboda and T. Pajdla. "Matching in Catadioptric Images with Appropriate Windows and Outliers Removal," *CAIP'01*.

[26] B. Triggs, P.F. McLauchlan, R.I. Hartley and A. Fitgibbon, "Bundle adjustment – a modern synthesis," *Vision Algorithms: Theory and Practice, 2000*.

[27] Y. Wu, H. Zhu, Z. Hu and F. Wu, "Camera Calibration from the Quasi-affine Invariance of Two Parallel Circles," *ECCV'04*.

[28] Z. Zhang, R. Deriche, O. Faugeras and Q.T. Luong, "A Robust Technique for Matching Two Uncalibrated Images through the Recovery of the Unknown Epipolar Geometry," *AI*, vol. 78, pp. 87-119, 1995.