Automatic Structure and Motion using a Catadioptric Camera

Maxime Lhuillier
LASMEA, UMR CNRS 6602, Université Blaise Pascal, 63177 Aubière, France
Maxime.Lhuillier@univ-bpclermont.fr, maxime.lhuillier.free.fr

Abstract

Methods for the robust and automatic estimation of scene structure and camera motion from image sequences acquired by a catadioptric camera are described. A first estimate of the complete geometry is obtained robustly from a rough knowledge of the two angles which defines the field of view. This approach is in contrast to previous work, which required mirror parameterization and only calculated (until now) the geometry of image pairs. Second, the additional knowledge of the mirror shape is enforced in the estimation. Both steps have become tractable thanks to the introduction of bundle adjustments for central and non-central cameras.

Finally, the system is presented as a whole, and many long image sequences are automatically reconstructed to show the qualities of the approach.

1. Introduction

The automatic estimation of scene Structure and camera Motion from image sequence ("Structure from Motion", or SfM) acquired by a perspective camera (whether calibrated or not) taken by hand has been a very active topic [12, 5] during the last decade, and many successful systems now exist [2, 21, 19]. On the other hand, during the same period, a large amount of works has been conducted on the geometric study of omnidirectional cameras/systems, although real omnidirectional and automatic SfM results for long image sequences have not yet been published. These cameras have a wide field of view, like dioptric (respectively, catadioptric) systems composed of a fish-eye (respectively, a mirror) in front of a perspective camera [9]. This paper focuses on SfM using a catadioptric system.

Central and Non-Central Cameras An important role is played by central omnidirectional cameras, i.e. cameras such that all back-projected rays intersect a single point in space (left of Figure 1). These cameras are, up to an image distortion, equivalent to a perspective camera, and are the most studied and used. The central cameras inherit many properties [7] from the perspective cameras, and should also inherit SfM algorithms once they are calibrated [10, 20]. The other systems are non-central (right of Figure 1).

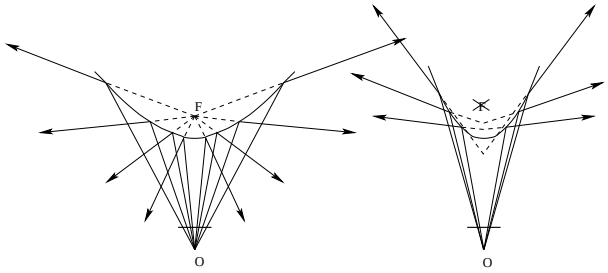

Figure 1: Examples of omnidirectional systems using aligned perspective cameras (with projection center O) and mirrors. Left: a central system is obtained, since all extended rays go through a single point F. Right: a non-central system.

Previous Work Closest works about central cameras are [13, 3, 8, 15]. In [13], calibration and successive essential matrices are estimated for a parabolic catadioptric camera (i.e. a parabolic mirror in front of an orthographic camera) from tracked points in the image sequence. Calibration initialization is given by the approximate field of view angle. Another 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. [3] estimate the motion of a calibrated omnidirectional camera using the essential matrix, and show that the results are better than the perspective camera motion when we look at objects far away, by taking advantage of the larger view field. The estimation of a more general central camera model is also proposed by [15], using direct modelization of the projection function and generalization of the simultaneous estimation of the usual fundamental matrix and one radial lens distortion coefficient [6]. Calibration parameters and successive essential matrices are estimated automatically using robust 9 and 15 points RANSAC algorithms and adequate bucketing technique [16].

Closest works about non-central catadioptric cameras are [1, 17]. Given a parameterization of the mirror [17], the non-central catadioptric camera is approximated by a central camera: the center is selected as that which minimizes the sum of squares of angular differences between real 3D rays reflected by the mirror surface to the scene point and the approximating rays emanating from the fictive center.

Once the model is simplified, the automatic method [15] is applied to obtain a first estimate of the image pair geometry. Finally, a Levenberg-Marquardt procedure is applied to refine the parameters of the two non-central catadioptric cameras. The error in the 3D space is minimized (instead of the usual re-projection errors in images) since the projection function of the non-central camera is not explicit. This is not ideal, since a such 3D error has no statistical foundation and 3D magnitude orders vary considerably between close and far away points. In a different context [1], the real-time pose estimation of a parabolic catadioptric system is achieved for planar motions in a room-size environment, using a few known 3D beacons. Pose accuracy is improved by replacing the ideal orthographic camera by a perspective one from the calibration step.

Our Work Section 2 describes our contributions for automatic SfM using a central catadioptric camera: geometry initialization and bundle adjustments. The geometry is estimated robustly from a rough knowledge of the two angles which define the field of view. Only 5 points RANSAC are needed in the geometry initialization of image pairs [18], without special bucketing techniques. This simple, very practical approach contrasts with the current, most advanced work [15, 17]. Section 3 describes our contributions to the non-central catadioptric camera: initialization from a first estimate with the central model, and bundle adjustments taking account of the mirror shape. In both cases, bundle adjustments minimize angle-based or image-based errors, and provide maximum likelihood estimate with the common (gaussian) noise assumption. Real bundle adjustments have not been previously designed and applied for catadioptric cameras, although it is well known [12] that they are the keystone for SfM, improving both the accuracy and robustness simultaneously. Experiments and conclusions are proposed in Section 4 and 5. We complete the system presentation in the Appendix with our matching method, and efficient calculations for non-central image projection and differentiation.

2. Central Catadioptric Camera

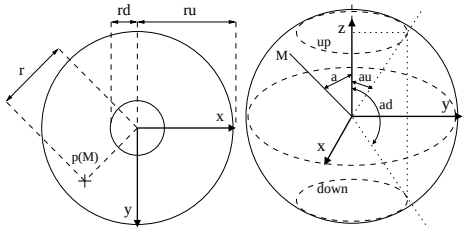

Figure 2: Left: two concentric circles of radii r_{up}, r_{down} (shortened ru, rd) in the retina plane. Right: angles $\alpha_{up}, \alpha_{down}, \alpha(X)$ (shortened au, ad, a) from the $z-$axis defi ne the fi eld of view for a point X by $\alpha_{up} \leq \alpha(X) \leq \alpha_{down}$. The circle "up" (respectively, "down") of rays on the right is projected by the central model on the big (respectively, small) circle on the left.

2.1. Camera Model and Notations

The central catadioptric model is defined by its orientation R (a rotation), the center $t \in \Re^3$ ($\Re$ is the real numbers), both expressed in the world coordinate system, and a "calibration" function $C : \Re^3 \rightarrow \Re^2$. If X is the homogeneous world coordinates of a 3D point such that the last coordinate is positive, the direction of the ray from t to X in the camera coordinate system is given by $d = R^\top (I_3 \quad -t) X$. The sign constraint for X is necessary to choose a ray (half line) among the opposite rays defined by directions d and $-d$, which are projected on two different points in the omnidirectional image. The projection $p(X)$ of X in the retina plane is finally obtained by $p(X) = C(d)$. If X is in the infinite plane, there are two projections: $C(d)$ and $C(-d)$.

Our central model also has a symmetry around the z-axis of the camera coordinate system: the omnidirectional image is between two concentric circles, and C is such that there is a positive function r such that

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

We introduce the notations r_{up} and r_{down} for the radii of both circles, the angle $\alpha = \alpha(x, y, z) = \alpha(X)$ between the z-axis and the ray direction, and the two angles α_{up} and α_{down} which define the field of view. We have $r_{down} \leq r \leq r_{up}$ and $\alpha_{up} \leq \alpha \leq \alpha_{down}$ as shown in Figure 2. Angles $\alpha_{up}, \alpha_{down}$ are approximately known and given by the mirror manufacturer, although they depend on the unknown focal length of the perspective camera and the relative positions between camera and mirror. The function r will be described later.

This model requires the knowledge of the mapping from the image to the retina, which maps the frontiers (ellipses) of the omnidirectional image to the two concentric circles. The mapping is pre-calculated for each image from an automatic ellipse detection (contour detection, polygonization, RANSAC, Levenberg-Marquardt), and is applied implicitly throughout Section 2 by proceeding as if the omnidirectional images frontiers are the two concentric circles. A similar model was previously used for fisheye cameras [15].

2.2. Geometry Initialization

Once the frontier circles and point matching (described in the Appendix) between image pairs have been obtained, the full geometry of the sequence can be estimated. We calculate pair geometries given an approximate calibration to take advantage of the usual perspective SfM tools.

The calibration function r is defined by the linear function of the angle α such that $r(\alpha_{up}) = r_{up}$ and $r(\alpha_{down}) = r_{down}$. The circle radii r_{down}, r_{up} are already estimated, and the approximate field of view angles $\alpha_{down}, \alpha_{up}$ are given by the mirror manufacturer. Such a choice assumes a similar image (radial) resolution for all displayed ray directions, which is not uncommon in practice.

Once the approximate calibration C is given, a ray direction for each image point is computable by the (explicit) reciprocal function C^{-1}. Now, known methods [4, 18] are applied to obtain a first estimate of the essential matrices, camera motions and 3D reconstructed points between all consecutive image pairs, and the geometry of all consecutive image triples by the pose calculation of the third camera once matches in 3 views have been reconstructed by the two others. The geometry refinement by bundle adjustment for omnidirectional cameras is described in Section 2.3.

The geometry of a sequence is estimated using a hierarchical approach [12]: 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 bundle adjustment.

2.3. Bundle Adjustments

Bundle adjustment improves the estimations of the centers t_i, the orientations R_i of cameras i and the positions X_j of points j by the minimization of a score: the sum of reprojection errors with known matched points m_{ij}.

Image Error At first glance, the score in the omnidirectional image space might be

$$\sum_{i,j} ||C(R_i^\top \begin{pmatrix} I_3 & -t_i \end{pmatrix} X_j) - m_{ij}||^2$$

with $||.||$ the Euclidean norm. However, practical problems occur for points X_j near the infinite plane with this score: some of them might go across the infinite plane during certain Levenberg-Marquardt iterations (when the sign of X_j fourth coordinate is modified), and this severely degrades the score. In fact, the projection in the omnidirectional image of a point lying exactly in the infinite plane is not one point, but two points $C(X)$ and $C(-X)$ (aligned and separated by the concentric circle center), and only one of these two reprojected points is very close to the detected point. If no care is taken when a point in a fixed ray goes across the infinite plane, its projections changes dramatically from $C(X)$ to $C(-X)$, and the global score follows. For this reason, we choose a continuous score in the infinite plane:

$$\sum_{i,j} min_{\epsilon=\pm 1} ||C(\epsilon R_i^\top \begin{pmatrix} I_3 & -t_i \end{pmatrix} X_j) - m_{ij}||^2.$$

Angular Error Another possible score is the sum of squared angles between the full line defined by points X_j and t_i, and the full line defined by point t_i and direction $C^{-1}(m_{ij})$. Theses angles are those between $C^{-1}(m_{ij})$ and $R_i^\top \begin{pmatrix} I_3 & -t_i \end{pmatrix} X_j$ in the i-th camera coordinate system. Let R_{ij} be a rotation which maps $C^{-1}(m_{ij})$ to $\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$, and $\pi(x,y,z) = \begin{pmatrix} x/z & y/z \end{pmatrix}$. These angles are those between $R_{ij}R_i^\top \begin{pmatrix} I_3 & -t_i \end{pmatrix} X_j$ and $\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$. Our score is

$$\sum_{i,j} ||\pi(R_{ij}R_i^\top \begin{pmatrix} I_3 & -t_i \end{pmatrix} X_j)||^2$$

by approximating the angles by their tangents. This approximation is acceptable since the angles are small for the inlier points near the solution. We also note that

- this angular score is independent of the R_{ij} choice
- the infinite plane problem mentioned for the image score does not occur here because $\pi(-x) = \pi(x)$
- both scores are valid without the X_j sign constraint.

3. Non-Central Catadioptric Camera

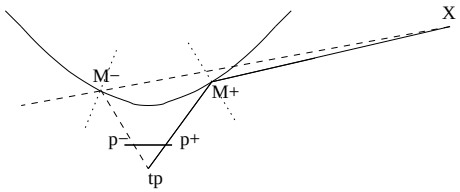

Figure 3: The ray which goes across point X and pinhole camera center t_p is reflected on the mirror at point $M^+(t_p, X)$ and projected by the pinhole camera to $p^+(X)$. We also define the point $M^-(t_p, X)$ on the mirror projected by the pinhole camera to $p^-(X)$, such that X is aligned with the reflected ray starting at $M^-(t_p, X)$ without being contained inside it. Points $M^+(t_p, X), p^+(X), M^-(t_p, X), p^-(X)$ are shortened M+,p+,M-,p- in this figure. Note that X,M-,M+ are not aligned.

3.1. Camera Model and Notations

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

The mirror surface contains the origin and has a symmetry around the z-axis of the mirror coordinate system. Let $M^+(A, B)$ be the function which gives the reflection point on the mirror (if any) for the ray which goes across points A and B. We have $A, B, M^+(A, B) \in \Re^3$, 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 the Appendix.

Let $\tilde{X}$ be the non-homogeneous world coordinates of a finite 3D point X. Since $R^\top(\tilde{X} - t)$ and $K^p R^{p\top} \begin{pmatrix} I_3 & -t^p \end{pmatrix}$ are respectively the X coordinates and the C^p matrix in the mirror coordinate system, we deduce that the homogeneous coordinates of the projection by the non-central camera of X are

$$p^+(X) \equiv K^p R^{p\top}(M^+(t^p, R^\top(\tilde{X} - t)) - t^p).$$

Figure 3 shows a cross section of the mirror and pinhole camera, the points $t^p, X, p^+(X), M^+(t^p, X)$.

Although this model does not directly define the projection for a point X_∞ in the infinite plane, we can define it by limit considerations. Let X be a finite point

which converges to X_∞ such that its homogeneous coordinates x, y, z are fixed and t converges to 0. Two limits of projections $p^+(X)$ are obtained by the relation $\tilde{X} = (\, x/t \quad y/t \quad z/t \,)$, one for each possible sign of t.

We also note that the parametrization (R^p, t^p, R) is not minimal because of the mirror symmetry around the z-axis: the expression of $p^+(X)$ is unchanged by replacing (R^p, t^p, R) by $(R_z R^p, R_z t^p, R R_z^\top)$ and applying the mirror symmetry relation $R_z M^+(A, B) = M^+(R_z A, R_z B)$ for any rotation R_z around the z-axis.

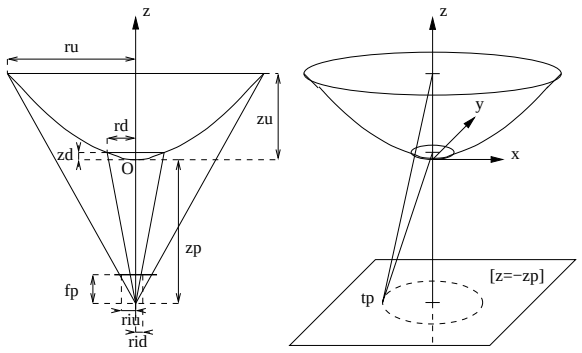

Figure 4: Left: useful (shortened) notations for z^p and f^p estimations. Right: t^p is located in a circle in the plane $z = -z^p$ given z^p and the angle between directions from t^p toward the centers of large and small border circles.

3.2. Geometry Initialization

Pinhole parameters The non-central parameters K^p, R^p, t^p are initialized from the two bounding ellipses detected in images and the mirror knowledge. These ellipses are the images of the known mirror border circles in z-planes $z = z_{up}$ and $z = z_{down}$ in the mirror coordinate system, with $z_{down} < z_{up}$.

The following realistic assumptions are made: the pinhole camera is pointing toward the mirror with t^p in the immediate neighborhood of mirror symmetry axis such that $R^p \approx I_3$, $t^p = (\, \epsilon_x^p \quad \epsilon_y^p \quad -z^p \,), 0 < z^p, |\epsilon_x^p| \ll z^p, |\epsilon_y^p| \ll z^p$, and the image ellipses are approximated by circles of radii r_{up}^i, r_{down}^i and centers c_{up}^i, c_{down}^i. Furthermore, we assume that K^p is the diagonal matrix with focal length f^p, focus point at image center and square pixels.

First, f_p and z_p may be estimated assuming $R^p = I_3$ and $\epsilon_x^p = \epsilon_y^p = 0$ by Thales relations $r_{up}^i / f^p = r_{up}/(z^p + z_{up})$ and $r_{down}^i / f^p = r_{down}/(z^p + z_{down})$ (left of Figure 4). The z^p estimation may also be obtained from the reflection law and knowledge of $r_{up}, z_{up}, \alpha_{up}$, or directly from the distance measurement between mirror and camera pinhole.

Second, $\epsilon_x^p, \epsilon_y^p$ are estimated. Since the angle between directions pointing toward the centers of mirror border circles is those between vectors $(K^p)^{-1} c_{up}^i$ and $(K^p)^{-1} c_{down}^i$, we know the radius of the circle in the plane $z = -z^p$ where t^p is located (right of Figure 4). Any t^p in this circle is possible by the over-parametrization (R^p, t^p, R). The hypothesis $R^p \approx I_3$ settles t^p as follow: from projection relations $\lambda_u \, (\, -\epsilon_x^p \quad -\epsilon_y^p \quad z^p + z_{up} \,)^\top = R^p \, (\, c_{up}^{i\top} \quad f^p \,)^\top$ and $\lambda_d \, (\, -\epsilon_x^p \quad -\epsilon_y^p \quad z^p + z_{down} \,)^\top = R^p \, (\, c_{down}^{i\top} \quad f^p \,)^\top$, we have $\lambda_u \approx \frac{f^p}{z_{up}+z^p} < \frac{f^p}{z_{down}+z^p} \approx \lambda_d$ and by subtraction $(\, c_{up}^{i\top} - c_{down}^{i\top} \quad 0 \,) \approx (\lambda_d - \lambda_u) \, (\, \epsilon_x^p \quad \epsilon_y^p \quad * \,)$. Thus we can find $\lambda > 0$ such that $(\, \epsilon_x^p \quad \epsilon_y^p \,)^\top = \lambda (c_{up}^i - c_{down}^i)$ from the circle radius in the plane $z = -z^p$. Finally, R^p is estimated from the two projections relations.

Mirror Pose and Scene Structure A first possibility is to identify notations t_i, R_i, X_j of both central and non central models, as soon as a complete and central geometry approximation is given by methods of Section 2. The scale factor, in fact, is not a free parameter as in the central case since it is fixed by the mirror size. We choose the initial scale factor for t_i and X_j such that the distances between consecutive mirrors are physically plausible compared with the mirror.

3.3. Bundle Adjustments

Bundle adjustment improves the estimations of the i-th mirror locations t_i and orientations R_i, the homogeneous coordinates $X_j = (\, X_j^x \quad X_j^y \quad X_j^z \quad X_j^t \,)$ of points j, by the minimization of a score: the sum of re-projection errors with known matched points m_{ij}. The explanations in this paragraph are very similar to those in Section 2.3.

Image Error At first glance, the score in the omnidirectional image space might be

$$\sum_{i,j} ||\pi(K_i^p R_i^{p\top} (M^+(t_i^p, R_i^\top (\tilde{X}_j - t_i)) - t_i^p)) - m_{ij}||^2$$

with $\tilde{X}_j = (\, X_j^x / X_j^t \quad X_j^y / X_j^t \quad X_j^z / X_j^t \,)$, the i-th pinhole parameters K_i^p, R_i^p, t_i^p, and $\pi(x, y, z) = (\, x/z \quad y/z \,)$. As in the central case, the image projection of a point X_j which goes across the infinite plane by a Levenberg-Marquardt iteration changes dramatically, and the resulting score ceases to be meaningful. The solution to this problem is the same for the non-central case: we redefine a score which is continuous for X_j in the neighborhood of the infinite plane.

In Figure 3 we introduce a kind of antipodal point for M^+ by a function M^- and minimize the score

$$\sum_{i,j} min_{\epsilon \in \{+,-\}} ||\pi(K_i^p R_i^{p\top} (M^\epsilon(t_i^p, R_i^\top (\tilde{X}_j - t_i)) - t_i^p)) - m_{ij}||^2.$$

In contrast to the simple choice of M^- such that $M^-(t^p, X) = M^+(t^p, -X)$, our choice is independent of the origin of mirror coordinates thanks to the relation $M^-(t^p, X) = M^+(t^p, 2M^-(t^p, X) - X)$. Both choices have same limits near the infinite plane and have very different image projection to that of M^+ elsewhere, as required.

Angular Error Let t_{ij} and n_{ij} be the i-th mirror point and the direction of the reflected ray defined by the image point m_{ij} and the i-th pinhole camera. A possible score is the sum of squared angles between the full line defined by the

points t_{ij} and X_j, and the full line defined by the point t_{ij} and the direction n_{ij}. These angles are those between n_{ij} and $R_i^\top \begin{pmatrix} I_3 & -t_i \end{pmatrix} X_j - t_{ij} X_j^t$ in the i-th mirror coordinate system. By introducing a rotation R_{ij} such that $R_{ij} n_{ij} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$ as in the central case, we obtain the score

$$\sum_{i,j} ||\pi(R_{ij}(R_i^\top \begin{pmatrix} I_3 & -t_i \end{pmatrix} X_j - t_{ij} X_j^t))||^2.$$

The three last remarks in Section 2.3 are still correct.

4. Experiments

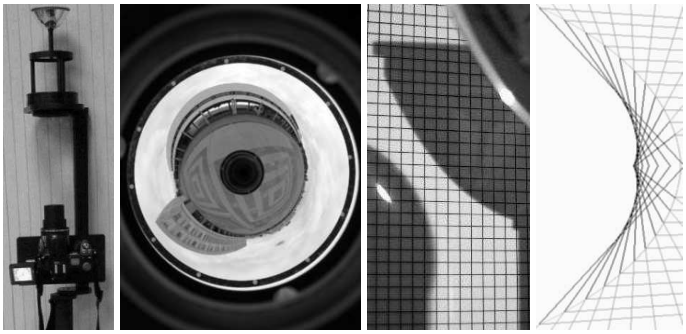

Figure 5: From left to right: the Kaidan "360 One VR" mirror with the Nikkon Coolpix 8700 camera (mounted on a monopod), a panoramic image (the view field is 360^0 horizontally and about 50^0 above and below the horizontal plane), the mirror shadow for profile estimation, and the mirror caustic for $z^p = 50cm$.

4.1. Our Context

Cheap Catadioptric Camera 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 symmetry axes of camera and mirror are not exactly the same (the axes alignment is manually adjusted and visually checked). According to the manufacturer (and the caustic size compared with the mirror), the system is not central.

Usual Camera Motions The camera motion described in the previous paragraph is close to a translation motion, for which the pinhole camera is pointing in a fixed direction (the sky). Such a motion is a critical motion for our central catadioptric model, as for the pinhole camera model (very similar proofs). Since many calibrations are possible for a given critical motion of the catadioptric camera, the full estimation of scene points and cameras is impossible. On the other hand, it would be easier to choose a prior calibration and help the SfM algorithms. The situation is not clear for our camera motions close to critical motions, and it should be examined experimentally for central/non central models.

4.2. Central Model

Panoramic images from omnidirectional images and top views of resulting reconstructions are shown in Figure 6.

Fountain The Fountain sequence is composed of 38 images of 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 movement. However, the systematic use of geometry refinement by bundle adjustment after each subsequence merging corrects all pair blunders, and the recovered camera motion is a smooth and circular motion around the fountain as expected.

Both unclosed and closed versions of this sequence are reconstructed. The unclosed one is obtained by duplication at the sequence end of the first image. The gap between both ends of the unclosed sequence is small: the distance between camera centers is 0.01 times the trajectory diameter, and the angle of relative orientation $R_{38} R_0^{-1}$ is 0.63^0.

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

House The House sequence is composed of 112 images taken in a cosy house, starting in the living room, crossing the lobby, having a loop in the kitchen, re-crossing the lobby and entering a bedroom.

Robustness and estimation stability for $\alpha_{up}, \alpha_{down}$ All central results above are obtained with a linear calibration function $r(\alpha)$ defined by the field of view angles $\alpha_{up} = 40^0, \alpha_{down} = 140^0$ given by the mirror manufacturer.

Two properties of the SfM method by the central model are explored in this paragraph: the robustness of initialization with very rough values of $\alpha_{up}, \alpha_{down}$, and the ability to refine these two angles using a cubic polynomial for $r(\alpha)$ (the four coefficients are new unknowns of the image bundle adjustment). Since the true camera motions are close but not exactly critical motions, both neither of the results for these two properties is certain.

Figure 7 shows good robustness of the hierarchical reconstruction: big inaccuracies of $\pm 20^0$ (respectively, $\pm 10^0$) degrees are usually (respectively, always) tolerated for $\alpha_{up}, \alpha_{down}$. Furthermore, the recovered $\hat{\alpha}_{up} = r^{-1}(r_{up})$ and $\hat{\alpha}_{down} = r^{-1}(r_{down})$ are reasonably stable for each sequence and physically plausible.

Some technical Details Computation times for 1632×1224-images with a P4 2.8GHz/800MHz are about 10s for each single image calculation (points and edge detections, circles), 20s for each image pair calculation (matching, essential matrix), and 5 min for the hierarchical reconstruction method applied to the House sequence (total time: 1 hour). About 700-1100 point matches satisfy the epipolar constraint between two consecutive images.

Figure 6: Top: panoramic images for image sequences. Bottom: top views of the recovered reconstructions with camera locations (little squares) and 3D points (black points). Left: Fountain (38 views, 5857 points). Middle: Road (54 views, 10178 points). Right: House (112 views, 15504 points). Some of these points are difficult to reconstruct when they are far 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.

	α_u	α_d	α_u	α_d	α_u	α_d	α_u	α_d	α_u	α_d
I	40	140	20	160	60	120	20	120	60	160
F	46	143	45	146	46	145	45	159	(46)	(144)
R	43	138	43	138	54	131	35	139	42	131
H	45	143	(45)	(143)	45	143	43	141	45	143

Figure 7: Each column pair α_u, α_d shows the field of view angles estimation for each of the 3 sequences Fountain, Road and House. These estimates are obtained from a SfM by the central model with initial angles defined in the row I. Brackets indicate a failure of the experiment, because the initial angles are too different from their exact values. The values given in brackets indicate the success of other experiments obtained with initial angles two times closer to the manufacturer angles than those in row I.

All bundle adjustments are implemented using analytical derivatives (see Appendix for the image error in the non-central case) and sparse calculations. The central angular and image errors are respectively used for the hierarchical reconstructions and $\hat{\alpha}_{up}, \hat{\alpha}_{down}$ estimations.

4.3. Non-Central Model

Non-central methods are also applied to enforce the mirror knowledge in the geometry estimation. The mirror profile function is a quartic, which is estimated from the rectification of the shadow shown in Figure 5. Once the non-central initialization is performed from the central results by measuring $z^p = 48cm$, the angular-based bundle adjustment is first applied because it is faster and more robust to initial conditions. Second, the image-based bundle adjustment is applied (as in the central case) since it provides a maximum likelihood estimation assuming that the image points are normally distributed around their true locations.

Real Images Non-central reconstructions for Fountain, Road and House are qualitatively similar as the central ones in Figure 6 (also see Figure 8). We also try many pinhole refinements with some further parameters in a final and image-error bundle adjustment, enforcing constraints such as fixed, common or independent focal length/orientation/position for all cameras. We found that these parameters (and the 3D scale factor) are difficult to estimate precisely: the convergence is slow starting from many physically plausible values.

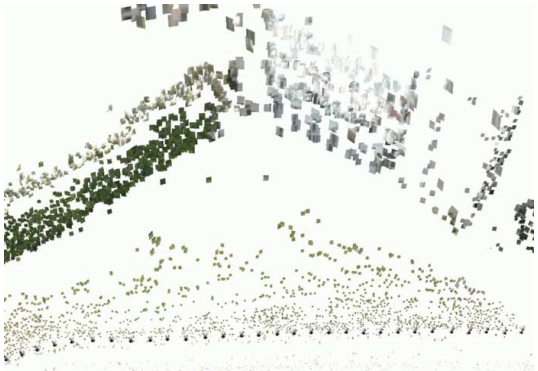

Figure 8: A local view of Road non-central reconstruction with a line of fir-trees (left), a building facade (right), and the ground. Patch size is proportional to distance between patch and camera.

3D Scale factor Estimation Synthesis experiments confirms the difficulty to estimate the 3D scale factor, although

this estimation is theoretically possible with our non-central catadioptric system. The ground truth reconstructions are half turn of a Fountain-like scene with trajectory radius r_t, mirror orientations R_i perturbed around I_3, 20 cameras and 1000 points well distributed in 3D and 2D spaces. First, image projections are corrupted by a noise of $\sigma = 1$ pixel and all camera locations t_i and points $\tilde{X}_j$ are multiplied by an initial factor s_{ini}. Second, angular and image-based non-central bundle adjustments are successively applied to these perturbed reconstructions, with known pinhole parameters ($z_p = 48cm$) and taking into account outliers. Table 1 shows the resulting RMS error (in pixels) and ratio s_{rec} between the recovered scale factor and its exact value. We note that the RMS has no clear minimum, and that the scale factor improvement are better for a small radius r_t.

	$r_t = 25cm$					$r_t = 2.5m$				
s_{ini}	0.5	.707	1	1.41	2	0.5	.707	1	1.41	2
s_{rec}	.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 ratio s_{rec} between the recovered scale factor and its exact value is given by the non-central refinement of a reconstruction perturbed by an initial homothety s_{init} and image noise $\sigma = 1$ pixel. The best values of s_{rec} are obtained for the trajectory with the smallest radius $r_t = 25cm$ (the ideal result is $s_{rec} = 1$).

4.4. Comparisons Between Camera Models

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

Consistency Table 2 shows consistencies between many reconstructions and the multi-view matching for our sequences using many criterions. The four camera models are: central models with linear and cubic calibration functions $r(\alpha)$, non-central models without and with pinhole parameters optimized by bundle adjustment (same t^p and f^p for all cameras, distinct R^p). The non-central models have the best consistencies: improvements about 5% (sometimes 12% for the Road) are obtained for the numbers of 3D and 2D inliers, with slightly slower RMS scores. The number of parameters is the same for central and non-central models without calibration parameter refinements.

	K	c-3d	nc-3d	c-2d	nc-2d	c-rms	nc-rms
F	n	5857	6221	31028	33262	0.84	0.78
F	y	5940	6223	32058	33876	0.78	0.76
R	n	10178	11312	50225	56862	0.87	0.82
R	y	10706	11313	54451	57148	0.78	0.79
H	n	15504	16432	75100	80169	0.83	0.78
H	y	15777	16447	77595	80311	0.76	0.76

Table 2: Consistencies between estimated geometries and data (matches in images) are given for both central and non-central models, for each of the 3 sequences Fountain, Road and House. Column K specifies experiments with (y) or without (n) calibration refinements. Each column proposes a consistency measure: c-3d, c-2d, c-rms are respectively the number of successfully reconstructed 3d points, the number of 2d points which have an image error less than 2 pixels, and the rms score. Similar notations nc-3d, nc-2d, nc-rms are used for the non-central models.

Difference 3D differences are given between central $(t_i^c, \tilde{X}_j^c)$ and non-central $(t_i^{nc}, \tilde{X}_j^{nc})$ models. The location and point differences are respectively $E_t^2 = \frac{1}{I}\sum_i ||S(t_i^c) - t_i^{nc}||^2$ and $E_x^2 = \frac{1}{J}\sum_j ||S(\tilde{X}_j^c) - \tilde{X}_j^{nc}||^2 / ||t_{i(j)}^{nc} - \tilde{X}_j^{nc}||^2$ with S the similarity transformation minimizing E_t, I the number of cameras, and J the number of 3D points. We obtain $E_t = 0.79cm, E_x = 0.027$ for the Fountain (respectively, $E_t = 2.6cm, E_x = 0.017$ and $E_t = 2.75cm, E_x = 0.054$ for the Road and House). The approximate trajectory lengths are 16, 42 and 22 meters, respectively.

Pose accuracy A real sequence (see Figure 9) is taken in an indoor controlled environment: the motion of our system is measured on a rail, in a room of dimensions $7m \times 5m \times 3m$. The trajectory is a 1 meter long straight line by translation, with 6 equidistant and aligned poses. We estimate the location error $E_t^2 = \frac{1}{6}\sum_i ||S(t_i^c) - t_i^g||^2$ with S the similarity transformation minimizing E_t and t_i^g the ground truth for camera location, and obtain $E_t^c = 1.1, E_t^{nc} = 1.2$ millimeters. Both models provides similar and good accuracies for the pose estimation in this context.

Figure 9: A panoramic image of the "controlled" sequence.

5. Summary and Conclusions

The first methods for the automatic estimation of scene structure and camera motion from long image sequences using catadioptric cameras are described in this paper, by introducing the adequate bundle adjustments (image and angular errors) and geometry initialization schemes for both central and non-central models. Many experiments about initialization robustness, accuracy, and comparisons between models are given for our non-central catadioptric camera. In many cases, the central model is a good approximation. Our system is also described as a whole. Although the results are very promising for our future applications (view synthesis, localization ...), more information is needed for an accurate estimation of pinhole parameters and the 3D scale factor.

Appendix

Initial Point Matching The usual matching procedure of interest points using correlation can not be directly applied to the omnidirectional images for two reasons: the matching ambiguity due to the repetitive patterns or textures in one image, and the geometric distortions between matched patches of two images taken from different view points. To our knowledge, previously published matching methods deal only with one of these problems.

In the context of the usual camera motions (roughly, translation motions with the pinhole camera pointing toward the sky), we observe that a high proportion of the distortions is compensated for by image rotation around the circle center. The Harris point detector [11] 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 patterns, the following procedure is applied. First, we try to match the points of interest of an image with themselves. The result is a "reduced" list of points which stores 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 patterns are greatly reduced, but the current list of matches is very incomplete. Third, this list is completed thanks to a quasi-dense match propagation [14]: the majority of image pixels are progressively matched using a 2D-disparity gradient limit, and two interest points roughly satisfying the resulting correspondence mapping between both images are added to the list. Such a list of matched interest points is obtained without any epipolar constraint, and is adequate for our geometry estimation methods.

Estimation and Derivation of $M^+(A,B)$ The non-central image error minimization by Levenberg-Marquardt in Section 3.3 requires efficient estimation and differentiation of $M^+(A,B)$. This function gives the reflection point on the mirror surface for the ray which goes across points A and B. Once these computations are done at (t^p, X) with pinhole center t^p and scene point X, all derivatives of the projection $p^+(X)$ are analytical according to the Chain Rule. Calculations are very similar for $p^-(X)$.

The known mirror is defined by the cylindric parameterization $f(r,\theta) = (\, r\cos(\theta) \quad r\sin(\theta) \quad z(r)\,) \in \Re^3$. Given $A, B \in \Re^3$, we are looking for (r,θ) such that $f(r,\theta)$ is the reflection point $M^+(A,B)$ in the mirror. Thus we have

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

by the reflection law, with $\vec{n}(r,\theta)$ the mirror normal. Only 2 among these 3 equations are independent: we are looking for (r,θ) such that $g(r,\theta,A,B) = 0$ with g defined by

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

and $R(-\theta)$ is the rotation of angle $-\theta$ around the 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 $\mathcal{C}^1$-function which map (A,B) to (r,θ) and its derivation

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

using the notation $d_X Y$ for the Jacobian of function $Y(X)$ with respect to parameters X. The value and derivatives of $M^+(A,B) = f(r,\theta)$ are deduced by function composition and Chain Rule.

References

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

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

[3] P. Chang and M. Hebert, "Omnidirectional structure from motion," *OMNIVIS'00*.

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

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

[6] A.W. Fitzgibbon, "Simultaneous Linear Estimation of Multiple View Geometry and Lens Distortion," *CVPR'01*.

[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] C. Geyer, T. Pajdla and K. Daniilidis, "Courses on Omnidirectional Vision," *ICCV'03*.

[10] J. Gluckman and S.K. Nayar, "Ego-Motion and Omnidirectional Cameras," *ICCV'98*.

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

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

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

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

[15] B. Micusik and T. Pajdla, "Estimation of Omnidirectional Camera Model from Epipolar Geometry," *CVPR'03*.

[16] B. Micusik and T. Pajdla, "Omnidirectional Camera Model and Epipolar Geometry Estimation by Ransac with Bucketing," *SCIA'03*.

[17] B. Micusik and T. Pajdla, "Autocalibration and 3D Reconstruction with Non-Central Catadioptric Cameras," *CVPR'04*.

[18] D. Nister. "An effi cient solution to the fi ve-point relative pose problem," *CVPR'03*.

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

[20] T. Pajdla, T. Svoboda and V. Hlavac, "Epipolar geometry of central catadioptric cameras," *IJCV*, vol. 49, no. 1, pp. 23-37.

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

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