A Generic Error Model and its Application to Automatic 3D Modeling of Scenes using a Catadioptric Camera

Maxime Lhuillier
LASMEA UMR 6602 UBP/CNRS,
24 avenue des Landais, 63177 Aubière Cedex
Tel.: +33-(0)4-73407593
Fax: +33-(0)4-73407262
maxime.lhuillier@univ-bpclermont.fr

This paper is published at the International Journal of Computer Vision (IJCV), volume 91, number 2, pages 175-199 (DOI: 10.1007/s11263-010-0374-2), Springer. It was received on March, 30th 2009 and accepted on July, 23th 2010. It can be downloaded at http://www.springerlink.com/content/m82358214369.

Abstract

Recently, it was suggested that structure-from-motion be solved using generic tools which are exploitable for any kind of camera. The same challenge applies for the automatic reconstruction of 3D models from image sequences, which includes structure-from-motion. This article is a new step in this direction. First, a generic error model is introduced for central cameras. Second, this error model is systematically used in the 3D modeling process. The experiments are carried out in a context which has rarely been addressed until now: the automatic 3D modeling of scenes using a catadioptric camera.

1 Introduction

There are two contributions in this article: the introduction of generic covariance and its application to the automatic 3D modeling of scenes using a catadioptric camera. The former and the latter are respectively summarized (and compared with previous works) in Sections 1.2 and 1.4. Section 1.1 explains why generic covariance has been introduced and Section 1.3 discusses the choice of a catadioptric camera for a visualization application. Almost all summarized results in Section 1.2 and a part of those of Section 1.4 are new materials over previous conference versions [19] and [17] of this work.

1.1 Toward Scene Modeling using Generic Tools

The automatic reconstruction of photo-realistic 3D models of scenes from image sequences taken by a moving camera is still a very active field of research. Once the camera parameters of the image sequence are recovered by structure-from-motion, dense stereo, and stereo merge into a single 3D model, are successively applied. Currently, many 3D modeling systems exist for perspective cameras [27], catadioptric cameras [3, 17] and multi-camera rig [7, 2], sometimes with the help of additional information such as odometry. Even if the intrinsic parameters of the camera are unknown, the methods involved depend on a given camera model.

Recently, it was suggested that the first step (SfM or Structure-from-Motion) be solved using a generic camera model and generic tools which are exploitable for any kind of camera, and the same challenge applies for the complete 3D modeling process. The practical advantage is the ability to change one camera model for another (or to mix different cameras). Many generic tools are already available for structure-from-motion: estimation of the generalized essential matrix [26, 21], pose calculation [24], bundle adjustment [28, 16], generic camera calibration [10, 14], and even SfM system [23] (without self-calibration).

Dense stereo is the second step. It is recognized that this step is very difficult in practice for uncontrolled environments. This difficulty increases in the generic context since the use of the image projection function is prohibited (this function is specific to the kind of camera). Furthermore, the global epipolar constraint is unavailable since the camera may be non-central: there are no matched curves between two images such that all 3D points which project on one curve also project on the other curve. Optical flow methods [12] may be used since they do not need the global epipolar constraint. An other method is suggested in [19]: assume that epipolar constraints are locally available and apply pair-wise stereo methods [29] after local rectifications.

The third step is the following: once cameras and matches between image pairs are known, the 3D model of the scene is reconstructed using generic tools. This model is a list of textured triangles in 3D which approximates the visible part of the scene where the camera has moved. We have to reconstruct 3D points, approximate them by a mesh, and deal with matching errors (false negatives and false positives), depth discontinuities, and a wide range of accuracies for reconstructed points (due to close foreground and far background, or view-point selection).

Generic covariance is introduced in [19] as a tool to deal with all items of the third step.

1.2 Generic Covariance for Central Camera

Virtual covariance of a 3D point is defined for 3D modeling of scenes using a catadioptric camera [17]. Once camera parameters are estimated by structure-from-motion, many local 3D models along the image sequence are reconstructed by dense stereo. Then virtual covariance is used to select the points of local models which are retained in the final and global model.

Generic covariance of a 3D point is defined for all central (single view point) cameras to solve the 3D modeling problem [19]. Since this second covariance only depends on the 3D point and the successive positions of the camera, it is highly independent with regard to the kind of camera. Its use is also extended to other issues of 3D modeling (hole filling, surface topology decisions ...).

Note that the original name of generic covariance in [19] is "virtual covariance". Here we have modified it since we think that the new name is more adequate: generic covariance does not depend on the projection function. We only assume that the camera is central.

The first theoretical contribution of the current paper is the definition of generic covariance itself. Indeed, Section 2 explains why the previous definition in [19] is naive and how to obtain a thorough (mathematically sound) definition. The thorough and naive definitions provide the same final expression of generic covariance.

In short, the thorough definition for a point $\mathbf{p}$ and several ray origins $\mathbf{o}_i$ is the following. First, ray directions $\mathbf{d}_i$ corresponding to $\mathbf{p}$ and $\mathbf{o}_i$ are calculated, and cost function E is defined such that $E(\mathbf{x})$ is a sum of discrepancies between directions $\mathbf{x} - \mathbf{o}_i$ and $\mathbf{d}_i$. Thus, $\mathbf{p}$ is the minimizer of E. Then, error models (random vectors) are defined for $\mathbf{d}_i$ in the unit sphere. There is a first-order error propagation of these errors to the minimizer of E, and the generic covariance of $\mathbf{p}$ is defined as the covariance of this minimizer.

One of the interests of generic covariance is its simplicity. It only depends on $\mathbf{p}, \mathbf{o}_i$ and a scale factor σ_α.

The second theoretical contribution of the paper is a list of properties of generic covariance. Section 3 provides the links (1) between unit sphere error and image error, (2) between virtual covariance [17] and generic covariance, (3) between σ_α and ray intersection problems from image points. In case (2), we give a simple criterion to check if both covariances are equal. In case (3), we explain how to estimate σ_α. Section 4 provides asymptotic properties of the uncertainty derived from generic covariance. Here we call "uncertainty" the length of the major semi-axis of the ellipsoid defined by generic covariance and a given probability. We show that uncertainty of a point $\mathbf{p}$ increases as the square of the distance between point $\mathbf{p}$ and the ray origins $\{\mathbf{o}_i\}$. This is a generalization for all central cameras of a well known property for rectified stereo-rig: depth $z = \frac{Bf}{d}$ has uncertainty $\sigma_z = \frac{z^2}{Bf}\sigma_d$ using B=baseline, f=focal length, d=image disparity and σ_d=disparity uncertainty.

There is an other asymptotic property observed in experiments [19] and proven in Section 4. In the two view case, the ratio of generic uncertainty by distance between $\mathbf{p}$ and $\{\mathbf{o}_1, \mathbf{o}_2\}$ is the inverse of apical angle (up to a constant scale). The apical angle defined by a point $\mathbf{p}$ and two ray origins $\mathbf{o}_1$ and $\mathbf{o}_2$ is the angle between $\mathbf{p} - \mathbf{o}_1$ and $\mathbf{p} - \mathbf{o}_2$. There is also a generalization in the multi-view case. The ratio between our uncertainty and distance is called "reliability".

This relation between apical angle and generic uncertainty has consequences on many previous works. On the one hand, reliability (derived from generic covariance) is thresholded in [19] to reject reconstructed points in 3D models which are considered "unreliable". On the other hand, apical angles are also thresholded to reject points [6] for the same reason. A criterion for video sub-sampling (used to stabilize structure-from-motion) is also based on apical angles [35]. Now we see the link between these methods: heuristic methods based on apical angles are roughly equivalent to generic covariance based methods. The former are two-view based and need recipes for more than two views. The latter are intrinsically multi-view.

1.3 Catadioptric Camera for Visualization Application

Our target application is the automatic reconstruction of photo-realistic 3D models for walkthroughs in a complex scene. This is a long-term research problem in Computer Vision and Graphics. A minimal requirement for interactive walkthrough is the scene rendering in any view direction around the horizontal plane, when the viewer moves along the ground. This suggests a wide field of view for the given images, for which many kinds of camera are possible [5]: catadioptric cameras, fish-eyes, or systems of multi-cameras pointing in many directions. Since we would like to capture any scene where a pedestrian can go, the hardware involved should be hand-held/head-held and not cumbersome. A catadioptric camera is a good candidate for all these constraints.

The main drawback of this choice is the low resolution compared with a perspective camera for a given field of view. We compensate for this problem using still image sequences taken with an equiangular catadioptric camera. Still images are preferred to video images because of their better quality (resolution, noise). An equiangular camera has also been selected from among other catadioptric cameras, since it is designed to spread the resolution of view field well throughout the whole image. Today, such a catadioptric camera can be purchased on the web for a few hundred Euros using adequate mirrors [1].

These two choices (still images and equiangular camera) mainly have two consequences. First, a still image sequence requires some effort and patience: the user should alternate a step in the scene and a (non blurred) shot by pressing a button. Second, an equiangular catadioptric camera is not a central camera. A non-central model complicates involved methods, since the back-projected rays do not intersect a single point in space as the perspective model.

This paper shows that the results obtained are worthwhile with such a setup, and that a central approximation of equiangular camera is sufficient in many cases.

1.4 Scene Modeling using Catadioptric Camera

Our 3D modeling methods are essentially generic for central cameras thanks to the systematic use of generic covariance. We integrate them in a fully automatic reconstruction system for catadioptric image sequences and experiment on hundreds of images without precise calibration knowledge. The overall system and experiments themselves are practical contributions of this work. Indeed, fully automatic 3D modeling from catadioptric images has rarely been addressed until now (although this is a long-standing problem for perspective images). All previous catadioptric approaches have been limited to dense reconstruction for a few view points or use accurate calibration of the camera. Experiments in Section 7 include outdoor scene reconstructions (only indoor examples are provided in previous works [13, 3, 9]). Now, the system is summarized.

Structure-from-motion (SfM) is the first step. Although the principles are well known, optimal (including global bundle adjustment) and robust SfM systems are not so common if the only given data is a long image sequence acquired by a general catadioptric camera and imprecise knowledge about the calibration. Previous catadioptric SfM systems are [33, 22, 18]. The method published in [18] is used in this work. This method is preferred to the generic method [23] since it also estimates camera calibration and it has more successful automatic matching. Here we use the central camera model with a general radial function, which is estimated from an approximate knowledge of the two view field angles provided by the mirror manufacturer. Details are omitted here since they have been published before.

Dense stereo is the second step. Catadioptric images are reprojected on virtual cubes, then dense stereo is applied on parallel faces such that conjugated epipolar lines are parallel. This is a particular case of the dense stereo generic scheme referenced in Section 1.1 and [19]. There are at least two reasons for using virtual cubes [17] instead of virtual cylinders [13, 3, 9] or even the original catadioptric images. First, a large choice of dense stereo methods is available [29]. Second, this facilitates the 3D reconstruction of the scene ground by pre-rectification. Here we do not focus on a specific stereo method. For convenience, we use the quasi-dense propagation method [20], but other methods are possible: [13, 3] use multi-baseline stereo inspired by [25], [9] uses graph-cut method [15] followed by postprocessings.

The third step is the 3D model generation from camera poses and image matches using generic covariance. Section 5 describes how to obtain a local (view-centered) 3D model for a few images and discusses limitations. First, a reference image is segmented by a 2D mesh using gradient edges and color information. Second, points are reconstructed from image matches by ray intersection. Third, 2D triangles are back-projected in 3D to fit the reconstructed points. Generic covariance is useful here to weight the minimized scores, to define the connections between triangles in 3D, to fill holes, and to reject unreliable triangles. Once many local 3D models have been reconstructed along the image sequence, generic covariance is once again used to obtain the final and global 3D model by view point selection and redundancy reduction (Section 6). Note that view point selection is a key issue [17]: a 3D point of the scene may be reconstructed in several local models at very different accuracies, and local models with the worst accuracies must not be used to reconstruct this point.

2 Definition of Generic Covariance

This Section provides two generic covariance definitions for central camera. The first one [19] is given in Section 2.2. Then Section 2.3 explains (1) why this definition is "naive" and (2) how to obtain a "thorough" definition. Last, the three steps of the thorough definition are set out in Sections 2.4, 2.5 and 2.6.

2.1 Notations

Several notations are used throughout the paper. Different fonts are used for reals (eg. $0, 1, a, \sigma$), vectors (eg. $\mathbf{b}$), matrices (eg. $\mathtt{C}$) and functions (eg. $F(x), g(x)$). The real vector space of dimension k is $\mathbb{R}^k$. The identity matrix of dimension k is $\mathtt{I}_k$. Let $\mathbf{k}$ be the vector $\begin{pmatrix}0 & 0 & 1\end{pmatrix}^T$ and $\mathbf{0}_k$ be the null vector of $\mathbb{R}^k$.

The angle between two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ is $(\mathbf{a}, \mathbf{b}) \in [0, \pi]$ and the cross product is $\mathbf{a} \wedge \mathbf{b}$. The unit sphere (not ball) of $\mathbb{R}^3$ is $\mathbb{S}^2$. The set of rotations of $\mathbb{R}^3$ is $SO(3)$. Let π be the function

$$\pi(\begin{pmatrix}x & y & z\end{pmatrix}^T) = \begin{pmatrix}\frac{x}{z} & \frac{y}{z}\end{pmatrix}^T. \quad (1)$$

Notation $\mathbf{z} \sim \mathcal{N}(\bar{\mathbf{z}}, \mathtt{C}_\mathbf{z})$ means that $\mathbf{z}$ is a Gaussian vector which has mean $\bar{\mathbf{z}}$ and covariance $\mathtt{C}_\mathbf{z}$.

2.2 Naive Definition

Let $\mathbf{o}_i \in \mathbb{R}^3, i \in \{1, 2, \cdots, I\}$ be many ray origins and a point $\mathbf{p} \in \mathbb{R}^3 \setminus \{\mathbf{o}_i\}$ such that $\mathbf{p}$ and $\{\mathbf{o}_i\}$ are not collinear. We introduce directions $\mathbf{d}_i \in \mathbb{S}^2$ and choose rotations $\mathtt{R}_i \in SO(3)$ such that

$$\mathbf{d}_i = \frac{\mathbf{p} - \mathbf{o}_i}{||\mathbf{p} - \mathbf{o}_i||} \text{ and } \mathtt{R}_i \mathbf{d}_i = \mathbf{k}. \tag{2}$$

Point $\mathbf{p}$ is a minimizer of the cost function

$$E(\mathbf{x}) = \sum_{i=1}^{I} ||\alpha_i(\mathbf{x})||^2 \text{ with } \alpha_i(\mathbf{x}) = \pi(\mathtt{R}_i(\mathbf{x} - \mathbf{o}_i)) \tag{3}$$

since $E(\mathbf{p}) = 0$. Using notation $\mathbf{z} = \mathtt{R}_i(\mathbf{x} - \mathbf{o}_i) = \begin{pmatrix} x & y & z \end{pmatrix}^T$, we have

$$||\pi(\mathbf{z})||^2 = \frac{x^2 + y^2}{z^2} = \tan^2(\mathbf{k}, \mathbf{z}) = \tan^2(\mathbf{d}_i, \mathbf{x} - \mathbf{o}_i). \tag{4}$$

Thus, $E(\mathbf{x})$ is the sum of squares of tangents of angles $(\mathbf{d}_i, \mathbf{x} - \mathbf{o}_i)$. Note that $\mathbf{p}$ is the unique minimizer of E since $\mathbf{p}$ and $\{\mathbf{o}_i\}$ are not collinear.

Let $\sigma_\alpha > 0$. We assume [19] that the angle errors α_i follow independent, isotropic and identical Gaussian errors $\mathcal{N}(\mathbf{0}_2, \sigma_\alpha^2 \mathtt{I}_2)$. Let J be the Jacobian of the function $\mathbf{x} \mapsto \begin{pmatrix} \alpha_1^T(\mathbf{x}) & \cdots & \alpha_I^T(\mathbf{x}) \end{pmatrix}^T$. There is a first-order error propagation from the α_i to the minimizer $\mathbf{p}$ of E: $\mathbf{p}$ follows Gaussian error with covariance matrix

$$C(\mathbf{p}) = \sigma_\alpha^2 (J(\mathbf{p})^T J(\mathbf{p}))^{-1}. \tag{5}$$

The next step is to calculate $C(\mathbf{p})$. Let J_π and J_{α_i} be the Jacobians of π and α_i. The Chain rule provides

$$J_{\alpha_i}(\mathbf{p}) = J_\pi(||\mathbf{p} - \mathbf{o}_i||\mathbf{k})\mathtt{R}_i = \frac{1}{||\mathbf{p} - \mathbf{o}_i||} \mathtt{A}\mathtt{R}_i \tag{6}$$

using $\mathtt{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. Thanks to Eqs. 5 and 6,

$$\begin{aligned} C^{-1}(\mathbf{p}) &= \frac{1}{\sigma_\alpha^2} J(\mathbf{p})^T J(\mathbf{p}) \\ &= \frac{1}{\sigma_\alpha^2} \sum_{i=1}^{I} \frac{1}{||\mathbf{p} - \mathbf{o}_i||^2} \mathtt{R}_i^T \mathtt{A}^T \mathtt{A} \mathtt{R}_i \\ &= \frac{1}{\sigma_\alpha^2} \sum_{i=1}^{I} \frac{1}{||\mathbf{p} - \mathbf{o}_i||^2} \mathtt{R}_i^T (\mathtt{I}_3 - \mathbf{k}\mathbf{k}^T) \mathtt{R}_i (7) \end{aligned}$$

Since $\mathtt{R}_i^T \mathtt{R}_i = \mathtt{I}_3$ and $\mathtt{R}_i^T \mathbf{k} = \mathbf{d}_i$, we obtain a simple expression of the generic covariance matrix:

$$C(\mathbf{p}) = \sigma_\alpha^2 (\sum_{i=1}^{I} \frac{\mathtt{I}_3 - \mathbf{d}_i \mathbf{d}_i^\top}{||\mathbf{p} - \mathbf{o}_i||^2})^{-1}. \tag{8}$$

Note that E and $C(\mathbf{p})$ do not depend on the $\mathtt{R}_i$ choice, but α_i and J do.

2.3 Against the Naive Definition

The first-order error propagation (for Gaussian vector) is formulated like this [11]. Assume that ϕ is a $\mathcal{C}^1$ continuous function, J_ϕ is the Jacobian of ϕ, and $\mathbf{y} \sim \mathcal{N}(\mathbf{y}_0, \mathtt{C}_\mathbf{y})$. Up to the first order, $\mathbf{y}$ propagates to

$$\phi(\mathbf{y}) \sim \mathcal{N}(\phi(\mathbf{y}_0), \mathtt{C}_\phi) \text{ with } \mathtt{C}_\phi = J_\phi(\mathbf{y}_0) \mathtt{C}_\mathbf{y} J_\phi^T(\mathbf{y}_0). \tag{9}$$

In the naive definition, Eq. 5 is obtained by propagation as if there were a function ϕ which maps value of $\begin{pmatrix} \alpha_1^T & \cdots & \alpha_I^T \end{pmatrix}^T$ to the minimizer of E. However, the value of E is fixed by the values of all α_i. In this case, the minimizer of E and ϕ itself are not well defined.

A correct use of propagation is the following:

1. Choose an error model for vectors $\{\mathbf{d}_i, \mathbf{o}_i\}$
2. Estimate Jacobian of function ϕ which maps model $\mathbf{y}$ of $(\mathbf{d}_1, \mathbf{o}_1, \cdots, \mathbf{d}_I, \mathbf{o}_I)$ to the minimizer $\mathbf{p}$ of E
3. Define the generic covariance $C(\mathbf{p})$ as the covariance of $\phi(\mathbf{y})$ using Eq. 9.

Henceforth, the subject of Section 2 is the development of these steps. Section 2.4 presents an error model choice (step 1) such that Eq. 5 is still correct. So Eq. 8 is too. The main part is the estimation of J_ϕ (step 2) in Section 2.5. Section 2.6 describes the so-called "first-order error propagation" (step 3).

2.4 Error Model of Ray Directions $\mathbf{d}_i$ and Origins $\mathbf{o}_i$

An ideal error model $\mathbf{d}_i^*$ of $\mathbf{d}_i$ modelizes perturbations of $\mathbf{d}_i$ on the unit sphere $\mathbb{S}^2$ since $\mathbf{d}_i \in \mathbb{S}^2$. In this article, we simplify the problem by using a Gaussian error model $\mathbf{d}_i^*$ which approximately modelizes perturbations on $\mathbb{S}^2$: $\mathbf{d}_i^*$ modelizes isotropic perturbations on T, the tangent plane on $\mathbb{S}^2$ at $\mathbf{d}_i$.

Let $\{\mathbf{a}, \mathbf{b}\}$ be an orthonormal basis of T, $\sigma_\alpha > 0$, $\mathbf{n}_i \sim \mathcal{N}(\mathbf{0}_2, \mathtt{I}_2)$ and $\mathbf{d}_i^* = \mathbf{d}_i + \sigma_\alpha \begin{pmatrix}\mathbf{a} & \mathbf{b}\end{pmatrix} \mathbf{n}_i$. Then, propagation from $\mathbf{n}_i$ to $\mathbf{d}_i^*$ implies $\mathbf{d}_i^* \sim \mathcal{N}(\mathbf{d}_i, \mathtt{C}_i)$ and

$$\begin{aligned} \mathtt{C}_i &= \sigma_\alpha^2 \begin{pmatrix}\mathbf{a} & \mathbf{b}\end{pmatrix} \mathtt{I}_2 \begin{pmatrix}\mathbf{a} & \mathbf{b}\end{pmatrix}^T \\ &= \sigma_\alpha^2 \begin{pmatrix}\mathbf{a} & \mathbf{b} & \mathbf{d}_i\end{pmatrix} (\mathtt{I}_3 - \mathbf{k}\mathbf{k}^T) \begin{pmatrix}\mathbf{a} & \mathbf{b} & \mathbf{d}_i\end{pmatrix}^T \\ &= \sigma_\alpha^2 (\mathtt{I}_3 - \mathbf{d}_i \mathbf{d}_i^T). \end{aligned} \qquad (10)$$

Last we assume that the $\mathbf{d}_i^*$ are independent with the same scale σ_α and the $\mathbf{o}_i$ have no errors.

2.5 Estimation of Jacobian J_ϕ

Let R_i be a $\mathcal{C}^2$ continuous function from a neighborhood of $\mathbf{d}_i$ into $SO(3)$ such that

$$R_i(\mathbf{x})\mathbf{x} = ||\mathbf{x}||\mathbf{k}. \qquad (11)$$

Now we define the cost function

$$E^*(\mathbf{x}) = \sum_{i=1}^{I} ||\alpha_i^*(\mathbf{x})||^2, \alpha_i^*(\mathbf{x}) = \pi(R_i(\mathbf{d}_i^*)(\mathbf{x} - \mathbf{o}_i)) \qquad (12)$$

and its minimizer $\mathbf{p}^*$. The goal of Section 2.5 is the estimation of Jacobian J_ϕ of function ϕ, which maps $(\mathbf{d}_1^*, \mathbf{d}_2^*, \cdots, \mathbf{d}_I^*)$ to $\mathbf{p}^*$.

Note that E^* is the sum of squares of the tangent of $(\mathbf{d}_i^*, \mathbf{x} - \mathbf{o}_i)$. Thus E^* and $\mathbf{p}^*$ do not depend on the R_i choice, and we have $E^* = E$ and $\mathbf{p}^* = \mathbf{p}$ if $\forall i, \mathbf{d}_i^* = \mathbf{d}_i$. Furthermore, ϕ does not depend on error model of ray origins since this model is not defined in our case (Section 2.4). Lemma 1 provides the existence of R_i.

Lemma 1 *Assume that* $\mathbf{a}, \mathbf{b} \in \mathbb{S}^2$. *There is a* $\mathcal{C}^2$ *continuous function* $\mathbf{x} \mapsto \mathtt{R}_\mathbf{x}$ *from the half space* $\{\mathbf{x} \in \mathbb{R}^3, \mathbf{a}^T\mathbf{x} > 0\}$ *into* $SO(3)$ *such that* $\mathtt{R}_\mathbf{x}\mathbf{x} = ||\mathbf{x}||\mathbf{b}$.

Proof First, assume $\mathbf{a}^T\mathbf{b} = 0$ and $\mathbf{a}^T\mathbf{x} > 0$. Thus $\mathbf{x}$ is not parallel to $\mathbf{b}$ and $\mathbf{x} \wedge \mathbf{b} \neq 0$. Let $\mathbf{n} = \frac{\mathbf{x} \wedge \mathbf{b}}{||\mathbf{x} \wedge \mathbf{b}||}$. Since $\mathbf{x}$ and $||\mathbf{x}||\mathbf{b}$ have the same norm and are both orthogonal to $\mathbf{n}$, the rotation $\mathtt{R}_\mathbf{x}$ defined by axis direction $\mathbf{n}$ and angle $(\mathbf{x}, \mathbf{b})$ is such that $\mathtt{R}_\mathbf{x}\mathbf{x} = ||\mathbf{x}||\mathbf{b}$. Function $\mathbf{x} \mapsto \mathtt{R}_\mathbf{x}$ is $\mathcal{C}^2$ continuous since functions $\mathbf{x} \mapsto (\mathbf{n}, (\mathbf{x}, \mathbf{b}))$ and $(\mathbf{n}, (\mathbf{x}, \mathbf{b})) \mapsto \mathtt{R}_\mathbf{x}$ are $\mathcal{C}^2$ continuous.

Second, assume $\mathbf{a}^T\mathbf{b} \neq 0$. Let $\mathtt{R}_0$ be a rotation such that $\mathbf{a}^T\mathtt{R}_0\mathbf{b} = 0$. We use the previous scheme to obtain a function $\mathtt{R}_\mathbf{x}^{'}$ such that $\mathtt{R}_\mathbf{x}^{'}\mathbf{x} = ||\mathbf{x}||\mathtt{R}_0\mathbf{b}$ with $\mathbf{x}$ in the half space $\mathbf{a}^T\mathbf{x} > 0$. So function $\mathtt{R}_\mathbf{x} : \mathbf{x} \mapsto \mathtt{R}_0^T\mathtt{R}_\mathbf{x}^{'}$ is such that $\mathtt{R}_\mathbf{x}\mathbf{x} = ||\mathbf{x}||\mathbf{b}$ for all $\mathbf{x}$ in the half space $\mathbf{a}^T\mathbf{x} > 0$. □Thanks to this Lemma, we know that E^* is well defined if $\mathbf{d}_i^*$ is in the half space $H_i = \{\mathbf{x} \in \mathbb{R}^3, \mathbf{d}_i^T\mathbf{x} > 0\}$. The error model of Section 2.4 meets $\mathbf{d}_i^* \in H_i$.

Now, we rewrite Eq. 12 in a different form to simplify the J_ϕ estimation. Let

$$\mathbf{d}_i^* \in H_i, \mathbf{x} \in \mathbb{R}^3 \setminus \{\mathbf{o}_i\}, \mathbf{y}^T = \begin{pmatrix}(\mathbf{d}_1^*)^T & \cdots & (\mathbf{d}_I^*)^T\end{pmatrix} \qquad (13)$$

and F be the function

$$F(\mathbf{x}, \mathbf{y}) = \begin{pmatrix}\alpha_1^*(\mathbf{x})^T & \cdots & \alpha_I^*(\mathbf{x})^T\end{pmatrix}^T. \qquad (14)$$

We obtain

$$E^*(\mathbf{x}) = ||F(\mathbf{x}, \mathbf{y})||^2 \text{ and } \phi(\mathbf{y}) = \arg\min_{\mathbf{x}} E^*(\mathbf{x}). \qquad (15)$$

Note that the estimation of J_ϕ is not the same as the standard estimation encountered in 3D Vision. E.g. for bundle adjustment, we should estimate J_ϕ such that

$$\phi(\mathbf{y}) = \arg\min_{\mathbf{x}} ||F(\mathbf{x}) - \mathbf{y}||^2 \qquad (16)$$

with $\mathbf{x}$ the 3D parameters (cameras poses and scene points), $\mathbf{y}$ the points detected in images, and F the projection functions. Then we have

$$J_\phi \approx (\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}})^{-1} \frac{\partial F}{\partial \mathbf{x}}^T. \qquad (17)$$

Our case is more general and is solved by Lemma 2.

Lemma 2 *Let F be a $\mathcal{C}^2$ continuous function with full rank Jacobian $\frac{\partial F}{\partial \mathbf{x}}$. There is a $\mathcal{C}^1$ continuous function*

$$\phi(\mathbf{y}) = arg\min_{\mathbf{x}} ||F(\mathbf{x}, \mathbf{y})||^2 \tag{18}$$

such that

$$J_\phi \approx -(\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}})^{-1} \frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{y}} \tag{19}$$

with derivatives of F taken at $(\phi(\mathbf{y}), \mathbf{y})$. This approximation is exact equality if $F(\phi(\mathbf{y}), \mathbf{y}) = 0$.

Proof Lemma 2 is a particular case of Proposition 6.1 in [8], which also asserts that function ϕ locally exists and is $\mathcal{C}^1$ continuous.

Notations F_k, x_i and y_j are the k-th, i-th and j-th coordinates of function F and vectors $\mathbf{x}$ and $\mathbf{y}$. The coefficient at i-th row and j-th column of matrix $\mathtt{M}$ is $\mathtt{M}_{i,j}$. Second-order partial derivatives of

$$f(\mathbf{x}, \mathbf{y}) = \frac{1}{2}||F(\mathbf{x}, \mathbf{y})||^2 \tag{20}$$

are

$$\frac{\partial^2 f}{\partial x_i \partial y_j} = \sum_k \{ \frac{\partial F_k}{\partial x_i} \frac{\partial F_k}{\partial y_j} + \frac{\partial^2 F_k}{\partial x_i \partial y_j} F_k \}. \tag{21}$$

The Gauss-Newton approximation of this equation is

$$\frac{\partial^2 f}{\partial x_i \partial y_j} \approx \sum_k \frac{\partial F_k}{\partial x_i} \frac{\partial F_k}{\partial y_j} = (\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{y}})_{i,j}. \tag{22}$$

A similar approximation is

$$\frac{\partial^2 f}{\partial x_i \partial x_{i'}} \approx \sum_k \frac{\partial F_k}{\partial x_i} \frac{\partial F_k}{\partial x_{i'}} = (\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}})_{i,i'}. \tag{23}$$

These approximations are exact equalities if $F(\mathbf{x}, \mathbf{y}) = 0$. Since $\phi(\mathbf{y})$ is minimizer of $\mathbf{x} \mapsto f(\mathbf{x}, \mathbf{y})$, we have

$$\forall i, \frac{\partial f}{\partial x_i}(\phi(\mathbf{y}), \mathbf{y}) = 0. \tag{24}$$

Using Eqs. 24, 22 and 23, we deduce that $\forall i, \forall j$,

$$\begin{aligned} 0 &= \frac{\partial}{\partial y_j}(\mathbf{y} \mapsto \frac{\partial f}{\partial x_i}(\phi(\mathbf{y}), \mathbf{y})) \\ &= \frac{\partial^2 f}{\partial x_i \partial y_j} + \sum_{i'} (\frac{\partial^2 f}{\partial x_i \partial x_{i'}}) \frac{\partial \phi_{i'}}{\partial y_j} \\ &\approx (\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{y}})_{i,j} + \sum_{i'} (\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}})_{i,i'} (\frac{\partial \phi}{\partial \mathbf{y}})_{i',j} \\ &\approx (\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{y}} + (\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}}) \frac{\partial \phi}{\partial \mathbf{y}})_{i,j}. \end{aligned} \tag{25}$$

Thanks to the full rank of $\frac{\partial F}{\partial \mathbf{x}}$, matrix $(\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}})$ is invertible and we obtain the result. □

Last, the following Lemma explicits $\frac{\partial F}{\partial \mathbf{x}}$ and $\frac{\partial F}{\partial \mathbf{y}}$.

Lemma 3 *Let $\mathbf{y}_0^T = \begin{pmatrix} \mathbf{d}_1{}^T & \cdots & \mathbf{d}_I{}^T \end{pmatrix}$, $\mathtt{R}_i = R_i(\mathbf{d}_i)$ and assume $\mathbf{d}_i = \frac{\mathbf{p} - \mathbf{o}_i}{||\mathbf{p} - \mathbf{o}_i||}$. We have*

$$\frac{\partial F}{\partial \mathbf{x}} = \begin{pmatrix} \mathtt{A}_1 \\ \vdots \\ \mathtt{A}_I \end{pmatrix} \text{ and } \frac{\partial F}{\partial \mathbf{y}} = \begin{pmatrix} \mathtt{B}_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathtt{B}_I \end{pmatrix} \tag{26}$$

where

$$\mathtt{A}_i = \frac{\mathtt{AR}_i}{||\mathbf{p} - \mathbf{o}_i||}, \mathtt{B}_i = -\mathtt{AR}_i, \mathtt{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{27}$$

The derivatives of F are taken at $(\mathbf{x}, \mathbf{y}) = (\mathbf{p}, \mathbf{y}_0)$.

Proof The block-wise structures of F derivatives result from the definition of F. First we calculate $\frac{\partial \alpha_i^*}{\partial \mathbf{x}}$ at $(\mathbf{x}, \mathbf{d}_i^*) = (\mathbf{p}, \mathbf{d}_i)$. Thanks to Eq. 11 and $\mathbf{d}_i = \frac{\mathbf{p} - \mathbf{o}_i}{||\mathbf{p} - \mathbf{o}_i||}$,

$$R_i(\mathbf{d}_i)(\mathbf{p} - \mathbf{o}_i) = ||\mathbf{p} - \mathbf{o}_i||\mathbf{k}. \tag{28}$$

We also need the Jacobian of π (π is defined in Eq. 1)

$$J_\pi(\begin{pmatrix} x & y & z \end{pmatrix}^T) = \begin{pmatrix} \frac{1}{z} & 0 & -\frac{x}{z^2} \\ 0 & \frac{1}{z} & -\frac{y}{z^2} \end{pmatrix}. \tag{29}$$

Then we apply the Chain rule (as in Eq. 6)

$$\mathtt{A}_i = \frac{\partial \alpha_i^*}{\partial \mathbf{x}}(\mathbf{p}, \mathbf{d}_i) = J_\pi(||\mathbf{p} - \mathbf{o}_i||\mathbf{k})\mathtt{R}_i = \frac{\mathtt{AR}_i}{||\mathbf{p} - \mathbf{o}_i||}. \tag{30}$$

Second we calculate $\frac{\partial \alpha_i^*}{\partial \mathbf{d}_i^*}$ at $(\mathbf{x}, \mathbf{d}_i^*) = (\mathbf{p}, \mathbf{d}_i)$. Using $\mathbf{d}_i = \frac{\mathbf{p} - \mathbf{o}_i}{||\mathbf{p} - \mathbf{o}_i||}$, we have

$$\alpha_i^*(\mathbf{p}) = \pi(R_i(\mathbf{d}_i^*)(\mathbf{p} - \mathbf{o}_i)) = \pi(R_i(\mathbf{d}_i^*)\mathbf{d}_i). \tag{31}$$

The Chain rule and $R_i(\mathbf{d}_i)\mathbf{d}_i = \mathbf{k}$ provide

$$\begin{aligned} \frac{\partial\alpha_i^*}{\partial\mathbf{d}_i^*}(\mathbf{p},\mathbf{d}_i) &= J_\pi(\mathbf{k})\frac{\partial(R_i(\mathbf{d}_i^*)\mathbf{d}_i)}{\partial\mathbf{d}_i^*}(\mathbf{d}_i)\\ &= \mathtt{A}\frac{\partial(R_i(\mathbf{x})\mathbf{d}_i)}{\partial\mathbf{x}}(\mathbf{d}_i). \end{aligned} \tag{32}$$

Fortunately, an explicit expression of R_i is not needed to calculate $\frac{\partial\alpha_i^*}{\partial\mathbf{d}_i^*}(\mathbf{p},\mathbf{d}_i)$: we derivate Eq. 11 with respect to parameter $a \in \{x,y,z\}$ of $\mathbf{x} = \begin{pmatrix}x & y & z\end{pmatrix}^T$

$$\frac{\partial(R_i(\mathbf{x}))}{\partial a}\mathbf{x} + R_i(\mathbf{x})\frac{\partial\mathbf{x}}{\partial a} = \mathbf{k}\frac{\partial||\mathbf{x}||}{\partial a} \tag{33}$$

and we obtain at point $\mathbf{x} = \mathbf{d}_i$

$$\frac{\partial(R_i(\mathbf{x})\mathbf{d}_i)}{\partial\mathbf{x}}(\mathbf{d}_i) + \mathtt{R}_i = \mathbf{k}\mathbf{d}_i^T. \tag{34}$$

We deduce from Eqs. 32 and 34 that

$$\mathtt{B}_i = \frac{\partial\alpha_i^*}{\partial\mathbf{d}_i^*}(\mathbf{p},\mathbf{d}_i) = \mathtt{A}(\mathbf{k}\mathbf{d}_i^T - \mathtt{R}_i) = -\mathtt{A}\mathtt{R}_i \tag{35}$$

□

Thanks to Lemma 3, $\frac{\partial F}{\partial\mathbf{x}}$ has full rank if $\mathbf{p}$ and $\{\mathbf{o}_i\}$ are not collinear. Then Lemma 2 is used: ϕ locally exists, it is $\mathcal{C}^1$ continuous, and its Jacobian at $(\mathbf{p},\mathbf{y}_0)$ is easy to estimate thanks to Eqs. 26 and 27.

2.6 First-Order Error Propagation

The last step of the generic covariance definition is the use of first-order error propagation (Eq. 9) with $\mathbf{y}^T = \begin{pmatrix}(\mathbf{d}_1^*)^T & \cdots & (\mathbf{d}_I^*)^T\end{pmatrix}$, ϕ of Section 2.5 and its Jacobian (Eq. 19):

$$\mathtt{C}_\phi = (\frac{\partial F}{\partial\mathbf{x}}^T\frac{\partial F}{\partial\mathbf{x}})^{-1}\frac{\partial F}{\partial\mathbf{x}}^T\frac{\partial F}{\partial\mathbf{y}}\mathtt{C}_\mathbf{y}\frac{\partial F}{\partial\mathbf{y}}^T\frac{\partial F}{\partial\mathbf{x}}(\frac{\partial F}{\partial\mathbf{x}}^T\frac{\partial F}{\partial\mathbf{x}})^{-1}. \tag{36}$$

at $(\mathbf{x},\mathbf{y}) = (\phi(\mathbf{y}_0),\mathbf{y}_0) = (\mathbf{p}, \begin{pmatrix}\mathbf{d}_1{}^T & \cdots & \mathbf{d}_I{}^T\end{pmatrix}^T)$.

Thanks to the $\mathbf{d}_i^*$ definition in Section 2.4, $\mathtt{C}_\mathbf{y}$ is a block-wise diagonal matrix with diagonal blocks equal to $\mathtt{C}_i$. Then Lemma 3 implies that $\frac{\partial F}{\partial\mathbf{y}}\mathtt{C}_\mathbf{y}\frac{\partial F}{\partial\mathbf{y}}^T$ is a block-diagonal matrix and its diagonal blocks are

$$\begin{aligned}\mathtt{B}_i\mathtt{C}_i\mathtt{B}_i^T &= \sigma_\alpha^2\mathtt{A}\mathtt{R}_i(\mathtt{I}_3 - \mathbf{d}_i\mathbf{d}_i^T)\mathtt{R}_i^T\mathtt{A}^T\\ &= \sigma_\alpha^2\mathtt{A}(\mathtt{I}_3 - \mathbf{k}\mathbf{k}^T)\mathtt{A}^T = \sigma_\alpha^2\mathtt{I}_2\end{aligned} \tag{37}$$

if $\mathbf{d}_i = \frac{\mathbf{p}-\mathbf{o}_i}{||\mathbf{p}-\mathbf{o}_i||}$. Thus, the generic covariance $C(\mathbf{p})$ is

$$\mathtt{C}_\phi = \sigma_\alpha^2(\frac{\partial F}{\partial\mathbf{x}}^T\frac{\partial F}{\partial\mathbf{x}})^{-1} = \sigma_\alpha^2(J(\mathbf{p})^TJ(\mathbf{p}))^{-1}. \tag{38}$$

This result is the same as that of the naive definition.

Observing Eq. 37, we note that alternative errors $\mathbf{d}_i^*$ are possible to obtain the same $C(\mathbf{p})$: conditions are $\mathtt{C}_i = \sigma_\alpha^2\mathtt{I}_3 + \lambda_i\mathbf{d}_i\mathbf{d}_i^T$, $\lambda_i \in \mathbb{R}$ ($\mathbf{d}_i$ is symmetry axis of $\mathtt{C}_i$). The alternative errors include non Gaussian errors $\mathbf{d}_i^* \in \mathbb{S}^2$, without the tangent plane approximation of Section 2.4.

3 Properties of Generic Covariance

First, Section 3.1 investigates how the $\mathbb{S}^2$ error (introduced in Section 2.4) is propagated to image space. Then Section 3.2 provides a criterion to check if the generic covariance propagated from $\mathbb{S}^2$ and the covariance propagated from image space are the same. In both cases, the input error is isotropic (in $\mathbb{S}^2$ or image) and has uniform scale (σ_α or σ_p). Last, Section 3.3 explains how to estimate σ_α.

3.1 Image Error and $\mathbb{S}^2$ Error

The error on the unit sphere $\mathbb{S}^2$ propagates to image error by the projection function of the central camera. This propagation is described in Lemma 4.

Lemma 4 *Let p be the projection function of the camera, and $\mathbf{d} \in \mathbb{S}^2$. Assume that p is $\mathcal{C}^1$ continuous with full rank Jacobian J_p. Up to the first order, the $\mathbb{S}^2$ error*

$$\mathbf{d}^* \sim \mathcal{N}(\mathbf{d},\mathtt{C}) \textit{ with } \mathtt{C} = \sigma_\alpha^2(\mathtt{I}_3 - \mathbf{d}\mathbf{d}^T). \tag{39}$$

propagates to image error

$$p(\mathbf{d}^*) \sim \mathcal{N}(p(\mathbf{d}),\mathtt{C}_p) \textit{ with } \mathtt{C}_p = \sigma_\alpha^2J_p(\mathbf{d})J_p^T(\mathbf{d}). \tag{40}$$

Furthermore, the image projection of the circular cone with infinitesimal aperture 2ϵ radians, apex $\mathbf{0}$, and axis $\mathbf{d}$ is in the ellipse

$$\{\mathbf{m} \in \mathbb{R}^2, (\mathbf{m} - p(\mathbf{d}))^T(\frac{\epsilon^2}{\sigma_\alpha^2}\mathtt{C}_p)^{-1}(\mathbf{m} - p(\mathbf{d})) = 1\}. \tag{41}$$

Proof Point $z\mathbf{d}$ defined by direction $\mathbf{d}$ and $z > 0$ is such that $p(z\mathbf{d}) - p(\mathbf{d}) = 0$ since $\mathbf{0}$ is the camera center. Using z and $\mathbf{d}$ derivatives of this equation, we obtain

$$J_p(z\mathbf{d})\mathbf{d} = 0 \text{ and } zJ_p(z\mathbf{d}) - J_p(\mathbf{d}) = 0. \tag{42}$$

Thus Eq. 39 propagates to $p(\mathbf{d}^*) \sim \mathcal{N}(p(\mathbf{d}), \mathtt{C}_p)$ with

$$\mathtt{C}_p = \sigma_\alpha^2 J_p(\mathbf{d})(\mathtt{I}_3 - \mathbf{d}\mathbf{d}^T)J_p^T(\mathbf{d}) = \sigma_\alpha^2 J_p(\mathbf{d})J_p^T(\mathbf{d}). \tag{43}$$

Now, we use SVD and introduce several notations:

$$J_p(\mathbf{d}) = \mathtt{UDV}^T,\ \mathtt{R} = \begin{pmatrix}\mathtt{V} & \mathbf{d}\end{pmatrix},\ \mathbf{x} = \mathtt{R}\begin{pmatrix}x & y & z\end{pmatrix}^T. \tag{44}$$

Matrix $\mathtt{R}$ is a rotation since $\mathbf{d} \in \mathbb{S}^2$ and $J_p(\mathbf{d})\mathbf{d} = \mathbf{0}$. Thus, $\epsilon^2 z^2 = x^2 + y^2$ iff $\mathbf{x}$ is in the circular cone with apex $\mathbf{0}$, axis $\mathbf{d}$ and aperture $2\arctan(\epsilon) \approx 2\epsilon$ radians. The linear Taylor expansion of p at $\mathbf{d}$ and Eq. 44 imply

$$\begin{aligned} p(\mathbf{x}) - p(\mathbf{d}) &= p(\frac{1}{z}\mathbf{x}) - p(\mathbf{d}) \approx J_p(\mathbf{d})(\frac{1}{z}\mathbf{x} - \mathbf{d}) \\ &\approx J_p(\mathbf{d})\mathtt{R}\begin{pmatrix}\frac{x}{z} & \frac{y}{z} & 0\end{pmatrix}^T = \frac{1}{z}\mathtt{UD}\begin{pmatrix}x \\ y\end{pmatrix} \end{aligned} \tag{45}$$

Note that Eq. 45 approximation is correct for $\mathbf{x}$ in the cone with small enough ϵ since $||\frac{1}{z}\mathbf{x} - \mathbf{d}||^2 = \frac{x^2+y^2}{z^2} = \epsilon^2$.

Furthermore, $\mathtt{D} > 0$ since J_p has full rank. Now, the SVD of $J_p(\mathbf{d})$, invertible $\mathtt{D}$, Eqs. 43 and 45 provide

$$\begin{aligned} &(p(\mathbf{x}) - p(\mathbf{d}))^T(\frac{\epsilon^2}{\sigma_\alpha^2}\mathtt{C}_p)^{-1}(p(\mathbf{x}) - p(\mathbf{d})) \\ &\approx \frac{1}{z^2}\begin{pmatrix}x & y\end{pmatrix}\mathtt{DU}^T(\epsilon^2\mathtt{UDV}^T\mathtt{VDU}^T)^{-1}\mathtt{UD}\begin{pmatrix}x & y\end{pmatrix}^T \\ &\approx \frac{x^2 + y^2}{z^2\epsilon^2}. \end{aligned} \tag{46}$$

We obtain the last result since $\epsilon^2 z^2 = x^2 + y^2$ iff $\mathbf{x}$ is in the circular cone. □

Lemma 4 also provides a method to visualize the covariance matrix $\mathtt{C}_p$ of the propagated image error $p(\mathbf{d}^*)$: uncertainty ellipse of image error $p(\mathbf{d}^*)$ is distortion ellipse of projection p centered at $p(\mathbf{d})$. We must remember that the distortion ellipse (or Tissot's indicatrix [34]) of p is the image projection by p of circular cone with infinitesimal aperture 2ϵ radians, apex $\mathbf{0}$ and axis $\mathbf{d}$.

Thus, there is a convenient and visual test to check if image error $p(\mathbf{d}^*)$ is "standard" (isotropic and uniform in the whole image): the more ellipses are circles with the same radius, the more standard the image error. Note that unavoidable distortions occur by local map from sphere into plane.

3.2 Generic and Virtual Covariances in 3D

On the one hand, the generic covariance $C(\mathbf{p})$ is defined by error propagation from the unit sphere $\mathbb{S}^2$ (Section 2). On the other hand, the virtual covariance $C_v(\mathbf{p})$ is defined by error propagation from the image [17]. Both are covariances for a 3D point $\mathbf{p}$. The former (the latter, respectively) assumes that the error amplitude in $\mathbb{S}^2$ (in the image, respectively) is uniform. The latter is specific to the projection function p of the camera. The following Lemma provides the condition to obtain $C_v = C$.

Lemma 5 *Let p be the projection function of the camera, $\mathtt{R}_i^0 \in SO(3)$, $\mathbf{o}_i \in \mathbb{R}^3$ and*

$$p_i(\mathbf{p}) = p(\mathtt{R}_i^0(\mathbf{p} - \mathbf{o}_i)). \tag{47}$$

Assume that p is $\mathcal{C}^1$ continuous. Let J_{p_i} be the Jacobian of $\mathbf{p} \mapsto p_i(\mathbf{p})$, $\sigma_p > 0$ and

$$C_v(\mathbf{p}) = \sigma_p^2(\sum_{i=1}^{I} J_{p_i}^T(\mathbf{p})J_{p_i}(\mathbf{p}))^{-1}. \tag{48}$$

The $\mathbb{S}^2$ error is

$$\forall \mathbf{d} \in \mathbb{S}^2, \mathbf{d}^* \sim \mathcal{N}(\mathbf{d}, \mathtt{C}) \text{ with } \mathtt{C} = \sigma_\alpha^2(\mathtt{I}_3 - \mathbf{d}\mathbf{d}^T). \tag{49}$$

If all $p(\mathbf{d}^*)$ have the same covariance $\sigma_p^2\mathtt{I}_2$, $C_v = C$.

Proof According to Lemma 4, $p(\mathbf{d}^*) \sim \mathcal{N}(p(\mathbf{d}), \mathtt{C}_p)$ with $\mathtt{C}_p = \sigma_\alpha^2 J_p(\mathbf{d})J_p^T(\mathbf{d})$. Furthermore, the SVD $J_p(\mathbf{d}) = \mathtt{UDV}^T$ and $\mathtt{C}_p = \sigma_p^2\mathtt{I}_2$ imply

$$\sigma_p^2\mathtt{I}_2 = \mathtt{C}_p = \sigma_\alpha^2\mathtt{UDV}^T\mathtt{VDU}^T \Rightarrow \mathtt{D} = \frac{\sigma_p}{\sigma_\alpha}\mathtt{I}_2. \tag{50}$$

Thus, there are $\mathtt{U}$, $\mathtt{V}$ and $\mathtt{R} = \begin{pmatrix}\mathtt{V} & \mathbf{d}\end{pmatrix}$ such that

$$\begin{aligned} & J_p(\mathbf{d}) = \frac{\sigma_p}{\sigma_\alpha}\mathtt{U}\mathtt{V}^T \text{ and } \mathtt{U}^T\mathtt{U} = \mathtt{I}_2, \mathtt{V}^T\mathbf{d} = \mathbf{0} \\ & \text{and } \mathtt{V}\mathtt{V}^T = \mathtt{R}(\mathtt{I}_3 - \mathbf{k}\mathbf{k}^T)\mathtt{R}^T = \mathtt{I}_3 - \mathbf{d}\mathbf{d}^T. \end{aligned} \tag{51}$$

Let $\mathbf{d}_i$ be such that $\mathbf{p} - \mathbf{o}_i = ||\mathbf{p} - \mathbf{o}_i||\mathbf{d}_i$. Using Chain Rule, Eqs 42 and 51, the Jacobian of p_i at $\mathbf{p}$ is

$$\begin{aligned} J_{p_i}(\mathbf{p}) &= J_p(\mathtt{R}_i^0(\mathbf{p} - \mathbf{o}_i))\mathtt{R}_i^0 = \frac{1}{||\mathbf{p} - \mathbf{o}_i||}J_p(\mathtt{R}_i^0\mathbf{d}_i)\mathtt{R}_i^0 \\ &= \frac{1}{||\mathbf{p} - \mathbf{o}_i||}\frac{\sigma_p}{\sigma_\alpha}\mathtt{U}_i\mathtt{V}_i^T\mathtt{R}_i^0 \end{aligned} \tag{52}$$

with

$$\mathtt{U}_i^T\mathtt{U}_i = \mathtt{I}_2, \mathtt{V}_i\mathtt{V}_i^T = \mathtt{I}_3 - \mathtt{R}_i^0\mathbf{d}_i(\mathtt{R}_i^0\mathbf{d}_i)^T. \tag{53}$$

Now, Eqs. 48, 52 and 53 imply

$$\begin{aligned} C_v^{-1}(\mathbf{p}) &= \frac{1}{\sigma_p^2}\sum_{i=1}^{I} J_{p_i}^T(\mathbf{p})J_{p_i}(\mathbf{p}) \\ &= \frac{1}{\sigma_p^2}\sum_{i=1}^{I}\frac{1}{||\mathbf{p} - \mathbf{o}_i||^2}\frac{\sigma_p^2}{\sigma_\alpha^2}\mathtt{R}_i^{0T}\mathtt{V}_i\mathtt{U}_i^T\mathtt{U}_i\mathtt{V}_i^T\mathtt{R}_i^0 \\ &= \frac{1}{\sigma_\alpha^2}\sum_{i=1}^{I}\frac{1}{||\mathbf{p} - \mathbf{o}_i||^2}(\mathtt{I}_3 - \mathbf{d}_i\mathbf{d}_i^T) \\ &= C^{-1}(\mathbf{p}). \end{aligned} \tag{54}$$

□

Lemma 5 asserts that generic and virtual covariances of a 3D point are the same if the image error is standard, i.e. if all $p(\mathbf{d}^*)$ have the same isotropic covariance. Thus, the visual test provided at the end of Section 3.1 can be used to check if generic and virtual covariances are the same.

In practice, perspective cameras with moderated field of view have distortion ellipses which are similar to circles with the same radius. According to Section 3.1, these cameras propagate the $\mathbb{S}^2$ error to the standard image error. According to Lemma 5, virtual and generic covariances are similar for perspective cameras.

3.3 Ray Intersection from Points in Images

Assume that we have a ray intersection problem: there are noisy and matched points in I images such that camera calibration and camera poses are known, and the corresponding 3D point $\mathbf{p}$ should be reconstructed. This problem may be solved as follows:

1. Calculate the ray directions $\mathbf{d}_i$ and origins $\mathbf{o}_i$ corresponding to image points (in world coordinates).
2. Choose rotations $\mathtt{R}_i$ such that $\mathtt{R}_i\mathbf{d}_i = \mathbf{k}$ and define the angle cost function $E(\mathbf{x})$ as in Eq. 3.
3. Estimate $\mathbf{p}$ as the minimizer of $E(\mathbf{x})$.

Here we can not define a covariance for $\mathbf{p}$ using Eq. 38 since it requires $\mathbf{d}_i = \frac{\mathbf{p}-\mathbf{o}_i}{||\mathbf{p}-\mathbf{o}_i||}$, which is wrong due to reconstruction in the presence of image noise. However, Eq. 36 does not need $\mathbf{d}_i = \frac{\mathbf{p}-\mathbf{o}_i}{||\mathbf{p}-\mathbf{o}_i||}$ and we use it to define covariance $C_r(\mathbf{p})$.

Now a 3D point $\mathbf{p}$ is reconstructed, we can reset $\mathbf{d}_i$ using $\mathbf{d}_i = \frac{\mathbf{p}-\mathbf{o}_i}{||\mathbf{p}-\mathbf{o}_i||}$ and define covariance $C(\mathbf{p})$ using Eq. 8. Both covariances C_r and C are obtained, the former comes from points detected in images and the latter comes from a reconstructed point in space.

Assume that the image noise is low, the mapping from image point to ray direction is continuous, and the mapping from ray directions to covariance is continuous by Eq. 36. Then C_r is approximated by C. This approximation is done in all our experiments.

Last we present a Lemma which is used in experiments to estimate σ_α from several ray intersection problems as defined above.

Lemma 6 *The mean (expected value) of random variable $E^*(\phi(\mathbf{y}))$ is*

$$\mathbb{E}(E^*(\phi(\mathbf{y}))) \approx (2I - 3)\sigma_\alpha^2. \tag{55}$$

Proof Here we need new notations $D(\mathbf{y}) = F(\phi(\mathbf{y}), \mathbf{y})$ and $\mathtt{P} = \frac{\partial F}{\partial \mathbf{x}}(\frac{\partial F}{\partial \mathbf{x}}^T\frac{\partial F}{\partial \mathbf{x}})^{-1}\frac{\partial F}{\partial \mathbf{x}}^T$. Note that $\mathtt{P}^2 = \mathtt{P} = \mathtt{P}^T$. Furthermore, Section 2.5 and Eq. 37 provide several relations which are useful for the current proof:

$$\mathbf{y}_0^T = \begin{pmatrix}\mathbf{d}_1^T & \cdots & \mathbf{d}_I^T\end{pmatrix}$$

$$\begin{array}{l}\phi(\mathbf{y}_0) = \mathbf{p},\ D(\mathbf{y}_0) = F(\mathbf{p}, \mathbf{y}_0) = 0 \\ \frac{\partial F}{\partial \mathbf{x}} J_\phi = -\mathtt{P}\frac{\partial F}{\partial \mathbf{y}},\ \frac{\partial F}{\partial \mathbf{y}} \mathtt{C}_\mathbf{y} \frac{\partial F}{\partial \mathbf{y}}^T = \sigma_\alpha^2 \mathtt{I}_{2I}. \end{array} \quad (56)$$

Partial derivatives of F are taken at $(\mathbf{x}, \mathbf{y}) = (\mathbf{p}, \mathbf{y}_0)$.

Now, a linear Taylor expansion of D is

$$\begin{aligned} D(\mathbf{y}) &\approx D(\mathbf{y}_0) + \frac{\partial D}{\partial \mathbf{y}}(\mathbf{y}_0).(\mathbf{y} - \mathbf{y}_0) \\ &\approx (\frac{\partial F}{\partial \mathbf{x}}(\mathbf{p}, \mathbf{y}_0) J_\phi(\mathbf{y}_0) + \frac{\partial F}{\partial \mathbf{y}}(\mathbf{p}, \mathbf{y}_0)).(\mathbf{y} - \mathbf{y}_0) \\ &\approx (\mathtt{I}_{2I} - \mathtt{P})\frac{\partial F}{\partial \mathbf{y}}(\mathbf{p}, \mathbf{y}_0).(\mathbf{y} - \mathbf{y}_0). \end{aligned} \quad (57)$$

Using first-order error propagation of $\mathbf{y} \sim \mathcal{N}(\mathbf{y}_0, \mathtt{C}_\mathbf{y})$, we have $D(\mathbf{y}) \sim \mathcal{N}(\mathbf{0}_{2I}, \mathtt{C}_D)$ such that

$$\begin{aligned} \mathtt{C}_D &= (\mathtt{I}_{2I} - \mathtt{P})\frac{\partial F}{\partial \mathbf{y}} \mathtt{C}_\mathbf{y} \frac{\partial F}{\partial \mathbf{y}}^T (\mathtt{I}_{2I} - \mathtt{P})^T \\ &= \sigma_\alpha^2(\mathtt{I}_{2I} - \mathtt{P} - \mathtt{P}^T + \mathtt{P}\mathtt{P}^T) = \sigma_\alpha^2(\mathtt{I}_{2I} - \mathtt{P}) \end{aligned} \quad (58)$$

Now the expected value of $E^*(\phi(\mathbf{y}))$ is

$$\begin{aligned} \mathbb{E}(E^*(\phi(\mathbf{y}))) &= \mathbb{E}(D(\mathbf{y})^T D(\mathbf{y})) \\ &= \mathbb{E}(\mathrm{tr}(D(\mathbf{y})D(\mathbf{y})^T)) \\ &= \mathrm{tr}(\mathbb{E}(D(\mathbf{y})D(\mathbf{y})^T)) \\ &\approx \mathrm{tr}(\mathtt{C}_D) = 2I\sigma_\alpha^2 - \sigma_\alpha^2 \mathrm{tr}(\mathtt{P}) \\ &= 2I\sigma_\alpha^2 - \sigma_\alpha^2 \mathrm{tr}((\frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}})^{-1} \frac{\partial F}{\partial \mathbf{x}}^T \frac{\partial F}{\partial \mathbf{x}}) \\ &= \sigma_\alpha^2(2I - 3). \end{aligned} \quad (59)$$

□

This lemma is used as follows. We reconstruct J points using the method described at the beginning of Section 3.3 for a same number of I images. Let E_j be the final value of the minimized score of the j-th point. Thanks to Eq. 55, we estimate σ_α using

$$(2I - 3)\sigma_\alpha^2 \approx \mathbb{E}(E^*(\phi(\mathbf{y}))) \approx \frac{1}{J}\sum_{j=1}^{J} E_j. \quad (60)$$

4 Properties of Generic Uncertainty

Generic uncertainty $U(\mathbf{p})$ is the length of the major semi-axis of the uncertainty ellipsoid defined by the generic covariance ($C(\mathbf{p})$ in Section 2) and a probability $p \in]0, 1[$. Let $\mathcal{X}_3^2(p)$ be the quantile function of the $\mathcal{X}^2$ distribution with 3 d.o.f. The ellipsoid is

$$\{\mathbf{x} \in \mathbb{R}^3, \quad (\mathbf{x} - \mathbf{p})^T C^{-1}(\mathbf{p})(\mathbf{x} - \mathbf{p}) \leq \mathcal{X}_3^2(p)\}. \quad (61)$$

We use notation $e(\mathbf{p})$ for the smallest singular value of $C^{-1}(\mathbf{p})$ and obtain

$$U(\mathbf{p}) = \sqrt{\frac{\mathcal{X}_3^2(p)}{e(\mathbf{p})}}. \quad (62)$$

We also define reliability [19]

$$R(\mathbf{p}) = \frac{U(\mathbf{p})}{\min_{i \in \{1, \cdots, I\}} ||\mathbf{p} - \mathbf{o}_i||} \quad (63)$$

The topic of this Section is the study of asymptotic properties of $U(\mathbf{p})$ and $R(\mathbf{p})$. By "asymptotic", we mean that $\mathbf{p}$ should be far enough from the ray origins $\{\mathbf{o}_i\}$.

These properties rely on the main result in Section 4.1, which expresses $e(\mathbf{p})$ as a function of a point in the convex hull CH of $\{\mathbf{o}_i\}$. Section 4.2 provides the link between apical angles and R, and confirms that $R(\mathbf{p})$ increases linearly as the distance between $\mathbf{p}$ and the ray origins $\{\mathbf{o}_i\}$. Thus, $U(\mathbf{p})$ increases quadratically.

4.1 Link Between $e(\mathbf{p})$ and the Convex Hull of $\{\mathbf{o}_i\}$

First, we need a Lemma whose assumptions are satisfied if point $\mathbf{p}$ is far enough from $\{\mathbf{o}_i\}$. The proof is very technical and may be skipped by the reader.

Lemma 7 *Let $\mathbf{p} \in \mathbb{R}^3$ such that*

$$\exists \mathbf{d}_0 \in \mathbb{S}^2, \forall i \in \{1..I\}, \mathbf{d}_i = \frac{\mathbf{p} - \mathbf{o}_i}{||\mathbf{p} - \mathbf{o}_i||}, \mathbf{d}_0^T \mathbf{d}_i \geq \frac{\sqrt{3}}{2}. \quad (64)$$

Let function G be

$$G : \mathbf{d} \in \mathbb{S}^2 \mapsto \sum_{i=1}^{I} w_i (\mathbf{d}_i^T \mathbf{d})^2 \textit{ with } w_i > 0. \quad (65)$$

There is a maximizer $\mathbf{d}$ of G such that $\forall i, \mathbf{d}_i^T \mathbf{d} \geq 0$.

Proof Since $(\mathbf{a},\mathbf{b})$ is the minimal path length between $\mathbf{a}$ and $\mathbf{b}$ on $\mathbb{S}^2$, we can use properties $(\mathbf{a},\mathbf{b}) \in [0,\pi]$, $\pi = (\mathbf{a},\mathbf{b}) + (-\mathbf{a},\mathbf{b})$, $(\mathbf{a},\mathbf{c}) \leq (\mathbf{a},\mathbf{b}) + (\mathbf{b},\mathbf{c})$ and $(\mathbf{a},\mathbf{b}) = (-\mathbf{a},-\mathbf{b})$. Furthermore, $\mathbf{d}_0^T\mathbf{d}_i \geq \frac{\sqrt{3}}{2}$ and $(\mathbf{d}_0,\mathbf{d}_i) \leq \frac{\pi}{6}$ are equivalent.

Let $\mathbf{d} \in \mathbb{S}^2$ such that $\frac{\pi}{3} < (\mathbf{d}_0,\mathbf{d})$. We have

$$\begin{aligned} & (\mathbf{d}_0,\mathbf{d}) \leq (\mathbf{d}_0,\mathbf{d}_i) + (\mathbf{d}_i,\mathbf{d}) \\ \Rightarrow\quad & \frac{\pi}{6} = \frac{\pi}{3} - \frac{\pi}{6} < (\mathbf{d}_0,\mathbf{d}) - (\mathbf{d}_0,\mathbf{d}_i) \leq (\mathbf{d}_i,\mathbf{d}) \end{aligned} \tag{66}$$

Let $\mathbf{d} \in \mathbb{S}^2$ such that $\frac{\pi}{3} < (-\mathbf{d}_0,\mathbf{d})$. We have

$$\begin{aligned} & (-\mathbf{d}_0,\mathbf{d}) \leq (-\mathbf{d}_0,-\mathbf{d}_i) + (-\mathbf{d}_i,\mathbf{d}) \\ \Rightarrow\quad & \frac{\pi}{6} = \frac{\pi}{3} - \frac{\pi}{6} < (-\mathbf{d}_0,\mathbf{d}) - (-\mathbf{d}_0,-\mathbf{d}_i) \leq (-\mathbf{d}_i,\mathbf{d}) \\ \Rightarrow\quad & (\mathbf{d}_i,\mathbf{d}) = \pi - (-\mathbf{d}_i,\mathbf{d}) < \pi - \frac{\pi}{6} = \frac{5\pi}{6}. \end{aligned} \tag{67}$$

Let $\mathbf{d} \in \mathbb{S}^2$. Thanks to Eqs. 66 and 67, we have

$$\begin{aligned} & \frac{\pi}{3} < (\mathbf{d}_0,\mathbf{d}) \text{ and } \frac{\pi}{3} < (-\mathbf{d}_0,\mathbf{d}) \\ \Rightarrow\quad & \frac{\pi}{6} < (\mathbf{d}_i,\mathbf{d}) < \frac{5\pi}{6} \Rightarrow |\mathbf{d}_i^T\mathbf{d}| < \frac{\sqrt{3}}{2} \\ \Rightarrow\quad & G(\mathbf{d}) = \sum_{i=1}^{I} w_i(\mathbf{d}_i^T\mathbf{d})^2 < \frac{3}{2}\sum_{i=1}^{I} w_i. \end{aligned} \tag{68}$$

However, $\mathbf{d}_0^T\mathbf{d}_i \geq \frac{\sqrt{3}}{2}$ and Eq. 68 imply

$$G(\mathbf{d}_0) = \sum_{i=1}^{I} w_i(\mathbf{d}_i^T\mathbf{d}_0)^2 \geq \frac{3}{2}\sum_{i=1}^{I} w_i > G(\mathbf{d}). \tag{69}$$

Now, we see that all $\mathbf{d} \in \mathbb{S}^2$ such that $\frac{\pi}{3} < (\mathbf{d}_0,\mathbf{d})$ and $\frac{\pi}{3} < (-\mathbf{d}_0,\mathbf{d})$ are not maximizers of G.

Let $\mathbf{d}$ be a maximizer of G ($\mathbf{d}$ is a singular vector of $\sum_{i=1}^{I} w_i\mathbf{d}_i\mathbf{d}_i^T$). We have $(\mathbf{d}_0,\mathbf{d}) \leq \frac{\pi}{3}$ or $(-\mathbf{d}_0,\mathbf{d}) \leq \frac{\pi}{3}$. If $(-\mathbf{d}_0,\mathbf{d}) \leq \frac{\pi}{3}$, $-\mathbf{d}$ is also a maximizer of G such that $(\mathbf{d}_0,-\mathbf{d}) = (-\mathbf{d}_0,\mathbf{d}) \leq \frac{\pi}{3}$. Thus, there is always a maximizer $\mathbf{d}$ of G such that $(\mathbf{d}_0,\mathbf{d}) \leq \frac{\pi}{3}$.

Last,

$$(\mathbf{d},\mathbf{d}_i) \leq (\mathbf{d},\mathbf{d}_0) + (\mathbf{d}_0,\mathbf{d}_i) \leq \frac{\pi}{3} + \frac{\pi}{6} = \frac{\pi}{2}. \tag{70}$$

This $\mathbf{d}$ is a maximizer of G such that $\forall i, \mathbf{d}_i^T\mathbf{d} \geq 0$. □

Here is the core of the asymptotic properties: the smallest singular value $e(\mathbf{p})$ of $C^{-1}(\mathbf{p})$ is a function of a point in the convex hull of the ray origins $\{\mathbf{o}_i\}$.

Theorem 1 *Let $\mathbf{p} \in \mathbb{R}^3$ such that*

$$\exists \mathbf{d}_0 \in \mathbb{S}^2, \forall i \in \{1..I\}, \mathbf{d}_i = \frac{\mathbf{p} - \mathbf{o}_i}{||\mathbf{p} - \mathbf{o}_i||}, \mathbf{d}_0^T\mathbf{d}_i \geq \frac{\sqrt{3}}{2}. \tag{71}$$

Let $e(\mathbf{p})$ be the smallest singular value of $C^{-1}(\mathbf{p})$ and

$$e(\mathbf{m},\mathbf{p}) = \frac{1}{\sigma_\alpha^2}\sum_{i=1}^{I}\frac{1}{||\mathbf{p} - \mathbf{o}_i||^2}(1 - (\mathbf{d}_i^T\frac{\mathbf{p} - \mathbf{m}}{||\mathbf{p} - \mathbf{m}||})^2). \tag{72}$$

There is $o(\mathbf{p})$ in the convex hull CH of $\{\mathbf{o}_i\}$ such that

$$e(\mathbf{p}) = \min_{\mathbf{m} \in CH} e(\mathbf{m},\mathbf{p}) = e(o(\mathbf{p}),\mathbf{p}). \tag{73}$$

Proof Thanks to the definition of $e(\mathbf{p})$ and the expression of $C(\mathbf{p})$ in Eq. 8,

$$\begin{aligned} e(\mathbf{p}) &= \min_{\mathbf{m} \in \mathbb{R}^3 \setminus \{\mathbf{p}\}} \frac{(\mathbf{m} - \mathbf{p})^T}{||\mathbf{m} - \mathbf{p}||} C^{-1}(\mathbf{p}) \frac{\mathbf{m} - \mathbf{p}}{||\mathbf{m} - \mathbf{p}||} \\ &= \min_{\mathbf{m} \in \mathbb{R}^3 \setminus \{\mathbf{p}\}} \frac{1}{\sigma_\alpha^2}\sum_{i=1}^{I}\frac{1}{||\mathbf{p} - \mathbf{o}_i||^2}(1 - (\mathbf{d}_i^T\frac{\mathbf{p} - \mathbf{m}}{||\mathbf{p} - \mathbf{m}||})^2) \\ &= \min_{\mathbf{m} \in \mathbb{R}^3 \setminus \{\mathbf{p}\}} H(\mathbf{m}) \text{ with } H(\mathbf{m}) = e(\mathbf{m},\mathbf{p}). \end{aligned} \tag{74}$$

We also apply Lemma 7 to function

$$G : \mathbf{d} \in \mathbb{S}^2 \mapsto \sum_{i=1}^{I}\frac{(\mathbf{d}_i^T\mathbf{d})^2}{||\mathbf{p} - \mathbf{o}_i||^2} : \tag{75}$$

there is a maximizer $\mathbf{d}_G$ of G such that $\forall i, \mathbf{d}_G^T\mathbf{d}_i \geq 0$. Since the assertions

- $\mathbf{m}$ is a minimizer of H
- $\frac{\mathbf{p} - \mathbf{m}}{||\mathbf{p} - \mathbf{m}||}$ is a maximizer of G,

are equivalent, we can choose a minimizer $\mathbf{m}$ of H using $\frac{\mathbf{p} - \mathbf{m}}{||\mathbf{p} - \mathbf{m}||} = \mathbf{d}_G$. Furthermore, we have

$$0 \leq \mathbf{d}_G^T\mathbf{d}_i = f_i(\mathbf{m}) \text{ with } f_i(\mathbf{x}) = \mathbf{d}_i^T\frac{\mathbf{p} - \mathbf{x}}{||\mathbf{p} - \mathbf{x}||}. \tag{76}$$

There is at least one $f_i(\mathbf{m}) > 0$ since the maximal value of G is not 0. We define the (no null) vector

$$d(\mathbf{m}) = \sum_{i=1}^{I}\frac{f_i(\mathbf{m})\mathbf{d}_i}{||\mathbf{p} - \mathbf{o}_i||^2}. \tag{77}$$

The Jacobian of H is

$$J_H(\mathbf{m}) = \frac{1}{\sigma_\alpha^2}\sum_{i=1}^{I}\frac{-2f_i(\mathbf{m})J_{f_i}(\mathbf{m})}{||\mathbf{p}-\mathbf{o}_i||^2} \tag{78}$$

with the Jacobian of f_i

$$J_{f_i}(\mathbf{m}) = \frac{\mathbf{d}_i^T(\mathbf{p}-\mathbf{m})}{||\mathbf{m}-\mathbf{p}||^3}(\mathbf{p}-\mathbf{m})^T - \frac{1}{||\mathbf{m}-\mathbf{p}||}\mathbf{d}_i^T. \tag{79}$$

Since $\mathbf{m}$ is a minimizer of H, Eqs. 78 and 79 imply

$$\begin{aligned} 0 &= -\frac{1}{2}\sigma_\alpha^2||\mathbf{m}-\mathbf{p}||J_H^T(\mathbf{m}) \\ &= \sum_{i=1}^{I}\frac{f_i(\mathbf{m})}{||\mathbf{p}-\mathbf{o}_i||^2}(\frac{\mathbf{d}_i^T(\mathbf{p}-\mathbf{m})}{||\mathbf{m}-\mathbf{p}||^2}(\mathbf{p}-\mathbf{m})-\mathbf{d}_i) \\ &= \frac{\mathbf{p}-\mathbf{m}}{||\mathbf{p}-\mathbf{m}||}\sum_{i=1}^{I}\frac{f_i^2(\mathbf{m})}{||\mathbf{p}-\mathbf{o}_i||^2} - \sum_{i=1}^{I}\frac{f_i(\mathbf{m})\mathbf{d}_i}{||\mathbf{p}-\mathbf{o}_i||^2}. \end{aligned} \tag{80}$$

Thanks to Eqs. 80 and 77 and $\mathbf{p}-\mathbf{m} = ||\mathbf{p}-\mathbf{m}||\mathbf{d}_G$,

$$\exists\lambda > 0, ||\mathbf{p}-\mathbf{m}||\mathbf{d}_G = \mathbf{p}-\mathbf{m} = \lambda d(\mathbf{m}). \tag{81}$$

Furthermore, $d(\mathbf{m})$ is a positive linear combination of $\mathbf{d}_i$ (thanks to Eqs. 77 and 76) and we obtain

$$\exists\lambda_i \geq 0, \mathbf{d}_G = \sum_{i=1}^{I}\lambda_i(\mathbf{p}-\mathbf{o}_i). \tag{82}$$

Point

$$\mathbf{o} = \mathbf{p} + \frac{-1}{\sum_{i=1}^{I}\lambda_i}\mathbf{d}_G = \frac{1}{\sum_{i=1}^{I}\lambda_i}\sum_{i=1}^{I}\lambda_i\mathbf{o}_i. \tag{83}$$

is both in the convex hull CH of $\{\mathbf{o}_i\}$ and in the line defined by direction $\mathbf{d}_G$ and point $\mathbf{p}$. Point $\mathbf{o}$ is a minimizer of H, as all other points in this line (except $\mathbf{p}$). Thanks to this definition of $\mathbf{o}$ and Eq. 74, we obtain

$$e(\mathbf{p}) = \min_{\mathbf{m}\in CH\setminus\{\mathbf{p}\}} e(\mathbf{m},\mathbf{p}) = e(o(\mathbf{p}),\mathbf{p}). \tag{84}$$

Last, Eq. 71 implies that $\mathbf{p} \notin CH$ and the proof is finished. □

4.2 Asymptotic Properties

First, we show in Eqs. 88 and 90 the relation between R and apical angles $(\mathbf{p}-\mathbf{o}_i, \mathbf{p}-\mathbf{o})$ or $(\mathbf{p}-\mathbf{o}_1, \mathbf{p}-\mathbf{o}_2)$. According to Theorem 1, there is $\mathbf{o} \in CH$ such that

$$e(\mathbf{p}) = \frac{1}{\sigma_\alpha^2}\sum_{i=1}^{I}\frac{1}{||\mathbf{p}-\mathbf{o}_i||^2}(1-(\mathbf{d}_i^T\frac{\mathbf{p}-\mathbf{o}}{||\mathbf{p}-\mathbf{o}||})^2). \tag{85}$$

Let $\beta_i = (\mathbf{p}-\mathbf{o}_i, \mathbf{p}-\mathbf{o})$. Since $\mathbf{p}$ is far from CH, $\sin\beta_i \approx \beta_i$ and $||\mathbf{p}-\mathbf{o}_i|| \approx ||\mathbf{p}-\mathbf{o}||$. Thus,

$$e(\mathbf{p}) = \frac{1}{\sigma_\alpha^2}\sum_{i=1}^{I}\frac{\sin^2\beta_i}{||\mathbf{p}-\mathbf{o}_i||^2} \approx \frac{1}{\sigma_\alpha^2||\mathbf{p}-\mathbf{o}||^2}\sum_{i=1}^{I}\beta_i^2. \tag{86}$$

Since

$$U(\mathbf{p}) = \sqrt{\frac{\mathcal{X}_3^2(p)}{e(\mathbf{p})}} \text{ and } R(\mathbf{p}) \approx \frac{U(\mathbf{p})}{||\mathbf{p}-\mathbf{o}||}, \tag{87}$$

we see that

$$R(\mathbf{p}) \approx \sigma_\alpha\sqrt{\frac{\mathcal{X}_3^2(p)}{\sum_{i=1}^{I}(\mathbf{p}-\mathbf{o}_i,\mathbf{p}-\mathbf{o})^2}}. \tag{88}$$

Similarly, if there are two ray origins, Theorem 1 implies

$$\begin{aligned} e(\mathbf{p}) &= \min_{\mathbf{o}\in[\mathbf{o}_1,\mathbf{o}_2]} e(\mathbf{o},\mathbf{p}) \\ &\approx \min_{|a|+|b|=(\mathbf{p}-\mathbf{o}_1,\mathbf{p}-\mathbf{o}_2)}\frac{a^2+b^2}{\sigma_\alpha^2||\mathbf{p}-\mathbf{o}||^2} \\ &\approx \frac{1}{\sigma_\alpha^2||\mathbf{p}-\mathbf{o}||^2}\frac{2(\mathbf{p}-\mathbf{o}_1,\mathbf{p}-\mathbf{o}_2)^2}{2^2} \\ &\approx \frac{(\mathbf{p}-\mathbf{o}_1,\mathbf{p}-\mathbf{o}_2)^2}{2\sigma_\alpha^2||\mathbf{p}-\mathbf{o}||^2} \end{aligned} \tag{89}$$

and we see that

$$R(\mathbf{p}) \approx \frac{\sigma_\alpha\sqrt{2\mathcal{X}_3^2(p)}}{(\mathbf{p}-\mathbf{o}_1,\mathbf{p}-\mathbf{o}_2)}. \tag{90}$$

Last, we show how U and R increase if $\mathbf{p}$ goes far from the ray origins $\mathbf{o}_i$. The double area of triangle $\mathbf{opo}_i$ has two expressions

$$||(\mathbf{o}-\mathbf{o}_i)\wedge(\mathbf{p}-\mathbf{o})|| = ||\mathbf{p}-\mathbf{o}_i||.||\mathbf{p}-\mathbf{o}||\sin\beta_i. \tag{91}$$

Since $\sin\beta_i \approx \beta_i$ and $||\mathbf{p}-\mathbf{o}_i|| \approx ||\mathbf{p}-\mathbf{o}||$, we obtain

$$(\mathbf{p}-\mathbf{o}_i, \mathbf{p}-\mathbf{o}) \approx \frac{1}{||\mathbf{p}-\mathbf{o}||}||(\mathbf{o}-\mathbf{o}_i) \wedge \frac{\mathbf{p}-\mathbf{o}}{||\mathbf{p}-\mathbf{o}||}||. \quad (92)$$

Eqs. 88 and 92 imply that $R(\mathbf{p})$ increases linearly as the distance between $\mathbf{p}$ and the ray origins. Thus, $U(\mathbf{p})$ increases quadratically.

5 Local 3D Models from Images

This Section explains how to obtain a local 3D model from a few images (I images). Local model reconstructs scene parts which are visible in a reference image.

The calibration function of the camera is known and maps image pixels to rays. It acts as a look-up table, as required in the generic context. A ray is a half line defined by its origin and direction. Thanks to the knowledge of camera pose by structure-from-motion (Section 1.4), origin $\mathbf{o}$ and direction $\mathbf{d}$ ($||\mathbf{d}|| = 1$) of rays are known in the world coordinate system.

First, the reference image is segmented by a 2D mesh using gradient edges and color information (Section 5.1). Second, 3D points are reconstructed by dense stereo and ray intersection (Section 5.2). Third, geometric and reliability tests are defined from generic covariance (Sections 5.3 and 5.4). Fourth, 2D triangles are back-projected in 3D to fit the reconstructed points by tacking into account depth discontinuities, hole filling, 3D point uncertainty and reliability (Section 5.5). Last, Section 5.6 discusses the extension of these methods for non-central cameras.

5.1 2D Mesh

Two standard assumptions are used in this step. First, the scene surface should be smooth enough to be approximated by a list of triangles in 3D. Second, the occluding contours and the tangent discontinuities of surfaces are projected at gradient edges or color discontinuities in images.

The 2D mesh in the reference image should satisfy many contradictory constraints: gradient edges at mesh edges, small enough mesh edges for good approximation of gradient edges, large enough mesh triangles for stable estimation of triangles in 3D, uniform sampling of the field of view, and good aspect ratio for triangles. A compromise is obtained with the method described in Appendix B.

The 2D mesh has two kinds of edges: constrained edges (image contours or image borders) and unconstrained edges (Delaunay edges).

5.2 Dense Stereo and Ray Intersection

In this step, dense multi-view correspondences are calculated and reconstructed.

We apply standard pair-wise stereo method after local rectifications (Section 1.1). In our context, the local rectifications are defined by mappings from catadioptric images into faces of virtual cubes. Then the quasi-dense propagation method [20] is applied between two parallel faces of two cubes. The resulting epipolar curves are conjugate and parallel lines, except for the faces which contains the epipole: the epipolar lines intersect the epipole at the face center. Virtual cubes are preferred to virtual cylinders and catadioptric (donut) images for the reasons given in Section 1.4. In practice, cube faces are slightly extended since 3D points may be projected in two cube faces which are not parallel.

The stereo method is defined for I images as follows. We consider each catadioptric image pair (*ref*, *sec*) with *ref* the reference image and *sec* a secondary image. First, two-view dense stereo is applied for a cube of *ref* and a cube of *sec* such that faces are pair-wise parallel. Second, the stereo results are combined in the original catadioptric image *ref*. For each pixel of *ref*, the corresponding points in all secondary images *sec* are obtained from the matching between cube faces and the mappings between cubes and catadioptric images. The calibration function provides ray origins $\mathbf{o}_i$ and directions $\mathbf{d}_i$ of matched points in images $i \in \{1, \cdots I\}$. Third, 3D points are estimated by the ray intersection method in Section 3.3. If the final value of cost function E is greater than a threshold (or if one of the I rays is not available), we can legitimately doubt the matching quality and no 3D point is retained. Small gaps (pixels of *ref* without 3D points) are filled with 3D points by interpolation.

5.3 Geometric Tests

Here we use the generic covariance $C(\mathbf{p})$ to define several tests, which are systematically used by mesh operations for 3D modeling. Eq. 8 provides the expression of $C(\mathbf{p})$ using point $\mathbf{p}$, ray origins $\{\mathbf{o}\}$ and scale σ_α. Section 3.3 describes a method to estimate σ_α.

Let Π be the plane $\mathbf{n}^\top \mathbf{x} + d = 0$. The Mahalanobis point-to-point and point-to-plane [30] squared distances are respectively

$$\begin{aligned} d^2(\mathbf{p}_1, \mathbf{p}_2) &= (\mathbf{p}_1 - \mathbf{p}_2)^\top C^{-1}(\mathbf{p}_1)(\mathbf{p}_1 - \mathbf{p}_2) \\ d^2(\mathbf{p}_1, \Pi) &= \min_{\mathbf{p}_2 \in \Pi} d^2(\mathbf{p}_1, \mathbf{p}_2) = \frac{(\mathbf{n}^\top \mathbf{p}_1 + d)^2}{\mathbf{n}^\top C(\mathbf{p}_1)\mathbf{n}} \end{aligned} \quad (93)$$

The point-to-point neighborhood test $T(\mathbf{p}_1, \mathbf{p}_2)$ is true if $d^2(\mathbf{p}_1, \mathbf{p}_2) \leq \mathcal{X}_3^2(p)$ and $d^2(\mathbf{p}_2, \mathbf{p}_1) \leq \mathcal{X}_3^2(p)$. Reminder: $\mathcal{X}_3^2(p)$ is defined in Eq. 61.

The point-to-plane neighborhood test $T(\mathbf{p}_1, \Pi)$ is true if $d^2(\mathbf{p}_1, \Pi) \leq \mathcal{X}_3^2(p)$.

The coplanarity test $T(\{\mathbf{p}_i\})$ is true if there is a plane Π such that all $T(\mathbf{p}_i, \Pi)$ are true. In practice, Π is estimated by random samples of 3 points in $\{\mathbf{p}_i\}$.

5.4 Reliability Test

A point $\mathbf{p}$ reconstructed from rays $(\mathbf{o}_i, \mathbf{d}_i), i \in \{1 \cdots I\}$ may be so inaccurate that the 3D model should not contain it. Such an example occurs if camera locations are collinear: $\mathbf{p}$ is inaccurate and should be rejected if it is too close to the line supporting the $\mathbf{o}_i$.

At first glance, we can decide that $\mathbf{p}$ is "reliable" enough for 3D modeling if uncertainty $U(\mathbf{p})$ (Eq. 62) is less than a threshold U_{max}. However, a reliability-based decision is preferred: the reliability test $T(\mathbf{p})$ is true if

$$R(\mathbf{p}) \leq R_{max} \text{ with } R(\mathbf{p}) = \frac{U(\mathbf{p})}{\min_i ||\mathbf{p} - \mathbf{o}_i||} \quad (94)$$

and R_{max} a threshold. Now we give advantages of reliability over uncertainty for the decision.

Since the camera is central, the reconstruction is defined up to a global 3D scale and a scale change of the whole reconstruction (3D points and camera centers) implies the same scale change of the uncertainties. Thus, uncertainty thresholding should have a threshold U_{max} which is proportional to the scene scale to obtain a decision which does not depend on the scale. A first reason to use Eq. 94 is its scale independence.

Furthermore, Eq. 94 allows reliable points for 3D modeling of the scene to have greater uncertainties if they are a long distance from ray origins $\mathbf{o}_i$ and smaller uncertainties if they are close. More precisely, the permitted maximal uncertainty is proportional to the distance between point $\mathbf{p}$ and ray origins $\mathbf{o}_i$. Thus, close foreground and far background of the scene are modelized at different uncertainties. This is better than the uncertainty thresholding $U(\mathbf{p}) < U_{max}$ which rejects too much background.

We note that the reliability test in Eq. 94 has two properties: (1) points in the neighborhood of the line supporting the $\mathbf{o}_i$ (if any) are unreliable and (2) the set of reliable points is bounded. These properties result from Eqs. 88 and 92.

5.5 2.5D Mesh

Now the geometric and reliability tests (Sections 5.3 and 5.4) are used to generate a 2.5D mesh by back-projection of the 2D mesh in the reference image using the dense cloud of reconstructed points. Sections 5.1 and Section 5.2 explain how to obtain the 2D mesh and the cloud, respectively.

The principle of the method is the following. First, the 2.5D mesh is initialized as a list of fully disconnected triangles in 3D. Then, this mesh is refined by alternating operations "Triangle Connection", "Hole Filling", "Triangle Removal", "Triangle Damping" and "Mesh Refinement". Last, unreliable triangles are rejected.

At any step, the 2.5D mesh in 3D is a back-projection of the 2D mesh in the reference image: each triangle t^{2d} of the 2D mesh corresponds to a triangle t^{3d} of the 2.5D mesh with vertices $\mathbf{v}_i \in \mathbb{R}^3$. The t^{3d} vertices are parametrized by depths $z_i > 0$ such that $\mathbf{v}_i = \mathbf{o}_i + z_i \mathbf{d}_i$ with $(\mathbf{o}_i, \mathbf{d}_i)$ the observation rays of t^{2d} vertices. Vertices in the 2D mesh may have many depths depending on current connections between triangles in 3D.

In Section 5.5, the index i is used for vertex numbering, not for image numbering as in all previous Sections. Thus $\forall i, \mathbf{o}_i = \mathbf{o}$ since the camera is central.

Mesh Initialization A RANSAC procedure is applied to each triangle t^{2d}. First, all 3D points reconstructed from pixels inside t^{2d} are collected in a list $L_{t^{2d}}$. Second, planes are calculated for random samples of 3 points in $L_{t^{2d}}$. Let Π be the plane minimizing

$$E^2_{t^{2d}}(\Pi) = \sum_{\mathbf{p} \in L_{t^{2d}}} \min\{\mathcal{X}^2_3(p), d^2(\mathbf{p}, \Pi)\} \tag{95}$$

with $\mathcal{X}^2_3(p)$ and $d(\mathbf{p}, \Pi)$ introduced in Eqs. 62 and 93. Then, we estimate depths z_i at the 3 vertices of t^{2d} such that $\mathbf{o}_i + z_i\mathbf{d}_i \in \Pi$ with $(\mathbf{o}_i, \mathbf{d}_i)$ the observation rays of these vertices. The triangle in 3D with three vertices $\mathbf{o}_i + z_i\mathbf{d}_i$ is added in the 2.5D mesh if $z_i > 0$.

Pair-Wise Triangle Connection Triangles in 3D should be connected to obtain a more realistic 3D model.

Let t^{3d}_a and t^{3d}_b be two 3D triangles such that the associated triangles t^{2d}_a and t^{2d}_b in the 2D mesh have a common edge (t^{2d}_a and t^{2d}_b are "weakly" connected). This edge has two vertices 0 and 1 in 2D, which correspond to triangle vertices $\{\mathbf{v}^a_0, \mathbf{v}^b_0\}$ and $\{\mathbf{v}^a_1, \mathbf{v}^b_1\}$ in 3D. The connection between t^{3d}_a and t^{3d}_b is made if the point-to-point neighborhood tests $T(\mathbf{v}^a_0, \mathbf{v}^b_0)$ and $T(\mathbf{v}^a_1, \mathbf{v}^b_1)$ defined in Section 5.3 are true.

The connection between t^{3d}_a and t^{3d}_b is defined as follows. Let z^a_i and z^b_i be depths such that $\mathbf{v}^a_i = \mathbf{o}_i + z^a_i\mathbf{d}_i$ and $\mathbf{v}^b_i = \mathbf{o}_i + z^b_i\mathbf{d}_i$ with $(\mathbf{o}_i, \mathbf{d}_i)$ the observation rays of 2D vertices $i \in \{0, 1\}$. New values of z^a_i and z^b_i are set to the former value of $\frac{1}{2}(z^a_i + z^b_i)$. Henceforth, the 2.5D mesh parameters z^a_i and z^b_i are linked by constraints $z^a_i = z^b_i$ for further processing.

Group-Wise Triangle Connection The "Pair-Wise Triangle Connection" above connects any triangle pair in 3D if they satisfy neighborhood conditions. Here we introduce the "Group-Wise Triangle Connection", which connects any k-group of triangles in 3D if they satisfy a coplanarity condition (typically $k \in \{2, 3, 4\}$).

A k-group of triangles in 3D is a list of k triangles t^{3d}_j such that the corresponding triangles t^{2d}_j are "strongly" connected in the 2D mesh. Two triangles are strongly connected if they have a common edge which is not constrained in the 2D mesh. We avoid constrained edges since they are potential surface discontinuities in 3D.

Any triangle pair $\{t^{3d}_a, t^{3d}_b\}$ in 3D is connected as in the pair-wise case if it is included in a k-group satisfying a coplanarity condition and if the corresponding $\{t^{2d}_a, t^{2d}_b\}$ in 2D have a common edge. Section 5.3 defines the coplanarity condition by $T(\{\mathbf{v}_i\})$ with $\{\mathbf{v}_i\}$ the list of all triangle vertices of the k-group.

Triangle Removal A smooth surface of the scene is expected to be approximated by a list of connected triangles in 3D. If a triangle is not connected to (at least) one of its neighbors after trials of triangle connections, we have some doubt as to its quality and may decide to remove it from the 2.5D mesh. They are many reasons for fully disconnected and bad triangles in 3D: false positive matches in images, triangle estimations using 3D points in both close foreground and far background, too few points for reliable estimation.

Triangle Damping The main drawback of "Triangle Removal" is the lack of triangles in scene parts which are not smooth such as tree foliage. If a triangle t^{3d} without connection is not removed, it may produce a major degradation of visual quality if it is very stretched in 3D in the direction $\mathbf{d}$ of ray which goes across t^{3d} center. In this case, the angle θ between t^{3d} normal $\mathbf{n}$ and $\mathbf{d}$ is greater than a threshold θ_0.

Thus, "Triangle Damping" reduces such degradations as follows: if $\theta_0 < \theta$, the t^{3d} depths z_i are disturbed such that (1) the t^{3d} center is fixed and (2) $\mathbf{n}$ is reset by $cos(\theta_0)\mathbf{d} + sin(\theta_0)\frac{\tilde{\mathbf{d}}}{||\tilde{\mathbf{d}}||}$ with $\tilde{\mathbf{d}} = \mathbf{n} - (\mathbf{n}^\top\mathbf{d})\mathbf{d}$. "Triangle Damping" may be preferred to "Triangle Removal" to obtain more triangles in the 3D model.

Hole Filling In our context, a hole is a connected component of triangles t_j^{2d} in the 2D mesh without corresponding triangles t_j^{3d} in the 2.5D mesh. "Hole Filling" is the definition of the lacking t_j^{3d} by interpolation of depths available in the hole border. Holes are mainly due to false negative matches in low textured areas. They degrade the visual quality of 3D model rendering if they are not properly filled.

The main risk is depth interpolation between foreground and background which also degrades the rendering quality, especially if foreground and background have different colors. We have the choice between strong connectivity (used in "Group-Wise Triangle Connection") and weak connectivity (used in "Pair-Wise Triangle Connection") between two triangles in the 2D mesh to define a hole as a connected component. The former is preferred to the latter, since the latter includes potential surface discontinuities at constrained edges too easily in the hole.

Thus, the hole border is a list of edges in the 2D mesh such that (1) edges are constrained or (2) edges are not constrained and have depths at their two vertices. All 3D points corresponding to these vertices with depths are collected in a list $\{\mathbf{v}_i\}$. We also define r as the ratio between the sum of 2D lengths of edges of type (2) and the sum of 2D lengths of all border edges.

We would like a well defined interpolation and a low risk of depth interpolation between foreground and background. Thus we request that the hole border is coplanar using coplanarity condition $T(\{\mathbf{v}_i\})$ defined in Section 5.3. We also request enough 3D information at the hole border using thresholding: $0.5 < r$.

If $T(\{\mathbf{v}_i\})$ is true, there is a plane Π which approximates the $\mathbf{v}_i$ and "Hole Filling" is defined as follows. Each vertex in the hole (including border) has a corresponding observation ray $(\mathbf{o}_i, \mathbf{d}_i)$ and a depth z_i defined by $\mathbf{o}_i + z_i\mathbf{d}_i \in \Pi$. Any hole triangle t^{2d} with positive z_i at its vertices defines a new triangle t^{3d} in the 2.5D mesh. We set depth constraints for further processing such that these vertices have only one depth.

Mesh Refinement The parameters of the 2.5D mesh is the list of depths z_i for each triangle vertex in 3D with many constraints (equalities) between the z_i. Operations "Hole Filling" and "Pair/Group-Wise Triangle Connection" are useful to increase the rendering quality of the 3D model, but they reduce the number of independent z_i and disturb the initial values of z_i obtained from the 3D point cloud. The consequence is an increasing discrepancy between the 3D point cloud and the 2.5D mesh. This problem is reduced by minimizing a global cost function including a discrepancy term and a smoothness term. The smoothness term is useful to reduce noise and enforce a prior knowledge of a piecewise smooth surface on the 2.5D mesh.

The cost function $e_{3d}(\{z_i\})$ is defined by

$$\sum_{t \in T} E_t^2 + \lambda \sum_{\{t_1,t_2\} \in Ed} \frac{1}{2}(|t_1| + |t_2|)(\mathbf{n}_{t_1} - \mathbf{n}_{t_2})^2 \quad (96)$$

with T the list of 2D mesh triangles which have triangles in 3D, $\{t, t_1, t_2\} \subset T$, $\{t_1, t_2\}$ the edge between triangles t_1 and t_2, $|t|$ the surface (in pixels) of t, Ed the list of unconstrained edges in the 2D mesh, and $\mathbf{n}_t$ the normal of the 3D triangle corresponding to the 2D triangle t. Weight λ is equal to 1 and E_t^2 is defined in Eq. 95. The cost function is minimized by a descent method with depths $\{z_i\}$ as parametrization. Depths have a wide range due to close foreground and far background, and this should be taken into account to reduce the cost efficiently. Let z_i^n be the i-th depth at iteration n of the descent method. At iteration $n+1$, we choose $z_i^{n+1} \in \{z_i^n - \delta_i(z_i^n), z_i^n, z_i^n + \delta_i(z_i^n)\}$ which minimizes the partial function $z_i \mapsto e_{3d}(z_i)$. Generic uncertainty is used to scale the increment δ_i by $\delta_i(z) = \epsilon U(\mathbf{o}_i + z\mathbf{d}_i)$ with $\epsilon = 0.02$.

Algorithm Summary Many combinations of the mesh operations above are possible and have been the subject of experiments. Our favorite strategy currently is

1. Mesh Initialization
2. Apply Group-Wise Triangle Connection ($k = 4$), Hole Filling and Mesh Refinement alternatively

3. Triangle Removal or Triangle Damping ($\theta_0 = \frac{7}{20}\pi$)

4. Apply Pair-Wise Triangle Connection, Hole Filling and Mesh Refinement alternatively.

5. Remove triangles with unreliable vertex (Eq. 94).

Step 2 connects triangles with strong conditions before step 3. Once step 3 has removed (or damped) improbable and unconnected triangles, step 4 connects triangles with weaker conditions.

5.6 From Central to 100% Generic Camera

The method in Section 5 has many limitations.

The first limitation is due to the use of a 2D mesh in a reference image (Section 5.1). The camera should not be too exotic to back-project connected image points (e.g. 2D triangles) to connected points in 3D (e.g. planar scene parts). Thus we should assume that the calibration function which maps pixels to rays in 3D is piecewise $\mathcal{C}^0$ continuous with known, smooth and polygonizable discontinuities. These discontinuities should be included in the 2D mesh border. Here is a simple example of discontinuity: the line between two composite images in the generic image of a stereo-rig.

The second limitation is due to the generic covariance definition used in the geometric and reliability tests (Sections 5.3 and 5.4). The definition of $C(\mathbf{p})$ requires the ray origins corresponding to $\mathbf{p}$. If the camera is (approximated by) a central camera or if $\mathbf{p}$ is reconstructed by ray intersection, the ray origins are known. They are unknown in other (non-central) cases, unless we apply the projection functions to $\mathbf{p}$ (but this is not a generic method). More investigations are needed to estimate efficiently ray origins in a generic camera framework for non-central cameras.

6 From Local to Global 3D Models

A local 3D model (Section 5) is not adequate for a complex scene since it is view-centered. Section 6 explains how to obtain a global 3D model of the scene from a list of local models reconstructed along the sequence. Typical lists are obtained by reconstruction of one local model for each I-tuple of consecutive still images (or video key-frames) of the sequence.

Global model reconstruction from local models involves view point selection, redundancy reduction, merging into topological manifold, mesh simplification, texture merging and packing. Our work only focuses on view point selection and redundancy reduction using generic covariance. Other topics are outside the paper scope.

6.1 View Point Selection

We formalize the view point selection problem as follows. Point $\mathbf{p}$ is reconstructed in local model l_0 and we would like to know if l_0 is one of the available local models which reconstruct $\mathbf{p}$ with the best accuracies. If so, $\mathbf{p}$ is retained in the global model. Note that $\mathbf{p}$ may be reconstructed in a large number of local models (in a typical list of local models) at very different accuracies, especially if the camera has a wide field of view.

Let $U_l(\mathbf{p})$ be the generic uncertainty (Eq. 62) of $\mathbf{p}$ using one ray origin $\mathbf{o}_i$ for each image of local model l. The list of reconstructed local models is L. Local model l_0 is one of the best local models to reconstruct $\mathbf{p}$ if

$$U^r_{l_0}(\mathbf{p}) \leq 1 + \epsilon \text{ with } U^r_{l_0}(\mathbf{p}) = \frac{U_{l_0}(\mathbf{p})}{\min_{l \in L} U_l(\mathbf{p})} \quad (97)$$

and threshold $\epsilon \geq 0$. In practice, $\epsilon > 0$ is useful to reduce the lack of triangles due to matching failures (false negatives).

At this point, several remarks arise. First, the view point selection defined by Eq. 97 requires the calculation of ray origins $\mathbf{o}_i$ for $\mathbf{p}$ and for all local models in L. As discussed in Section 5.6, this problem is not yet solved in the non-central case. This is not a problem in our case where the camera is (approximated by) a central camera: the $\mathbf{o}_i$ are the locations of camera center. Second, we note that Eq. 97 does not depend on probability p and scale σ_α involved in the generic uncertainty definition (Eqs. 62 and 5). Indeed, changing p or σ is a multiplication of all $U_l(\mathbf{p})$ by the same value.

A triangle of l_0 is retained in the global model if (at least) one of its three vertices $\mathbf{p}$ satisfies Eq. 97. Thus, the time complexity of the view point selection method is $|L|^2|V|$ with $|L|$ the number of local models and $|V|$ the number of 3D vertices in a local model ($|V|$ is assumed to be constant).

Note that this definition of view point selection does not depend on $\mathbf{p}$ visibility in views of the sequence (nor does the generic uncertainty definition). In fact, the use of $U_l(\mathbf{p})$ for view point selection has no sense if $\mathbf{p}$ can not be reconstructed by l due to lack of visibility. Thus Eq. 97 is improved by taking visibility into account as follows: we reset $U_l(\mathbf{p}) = +\infty$ if $\mathbf{p}$ is not in the view field of (at least) one image of l. We may also consider the global surface to be reconstructed as a possible occluder for $\mathbf{p}$ in each image of l, but this problem is not integrated in this paper.

6.2 Redundancy Reduction

The method in Section 6.1 generates a global model from selected parts of local models which have the smallest uncertainties. Although the result is ready for visualization, a redundancy reduction method is useful to decrease the number of triangles.

Triangles with the largest uncertainties are progressively removed if they are overlapped (up to their uncertainty) by other triangles of the global model. A region decreasing-like method is used. At each step, we focus on the triangle t with the largest generic uncertainty which is on the border of one of the meshes currently retained in the global model. The uncertainty volume of t is defined as follows. The t vertices are $\mathbf{v}_i = \mathbf{o}_i + z_i\mathbf{d}_i, i \in \{1, 2, 3\}$ with depths z_i and observations rays $(\mathbf{o}_i, \mathbf{d}_i)$ in the reference image of the local model l which reconstructs t. The volume is a truncated triangular cone with face $f_+ : \mathbf{v}_i + U_l(\mathbf{v}_i)\mathbf{d}_i$ and face $f_- : \mathbf{v}_i - U_l(\mathbf{v}_i)\mathbf{d}_i$. Then we test if t is overlapped by other triangles: (1) all triangles which intersect the volume are collected in a list (2) the volume is sub-sampled into segments connecting f_+ and f_- (3) t is overlapped if all segments intersect triangle(s) in the list. If t is overlapped, it is removed from the global model. The process stops when there is no overlapped triangle in mesh borders. In practice, step (1) is accelerated using hierarchical bounding boxes and test eliminations.

7 Experiments

First, Sections 7.1 and 7.2 illustrate properties of generic covariance (Sections 3 and 4). Second, experiments are provided for our catadioptric cameras. The accuracy of the local 3D model reconstruction method (Section 5) is evaluated on a synthetic scene in Section 7.3. Section 7.4 illustrates the view point selection method (Section 6) to generate the global 3D model. Last, the overall system is experimented in Section 7.5 on real image sequences.

7.1 Properties of Generic Uncertainty and Reliability

Figure 1 shows generic uncertainty U and reliability R using values $\mathcal{X}_2^3(0.9) = 6.25$ and $\sigma_\alpha = 0.001$. Black edges are the main axes of uncertainty ellipsoids centered at certain points $\mathbf{p}$ (their length is $2U(\mathbf{p})$). Every point $\mathbf{p}$ has a gray level depending on the interval where $R(\mathbf{p})$ lies: $[0, \frac{1}{40}[, [\frac{1}{40}, \frac{2}{40}[, [\frac{2}{40}, \frac{3}{40}[, [\frac{3}{40}, \frac{4}{40}[, [\frac{4}{40}, +\infty]$ (darkest gray levels for largest R).

In the first case (on the left), we have two ray origins $\mathbf{o}_1$ and $\mathbf{o}_2$. Both $U(\mathbf{p})$ and $R(\mathbf{p})$ are defined everywhere (except on the line defined by $\mathbf{o}_1$ and $\mathbf{o}_2$). Due to the symmetry of the problem, U and R are the same for any plane in 3D containing $\mathbf{o}_1$ and $\mathbf{o}_2$. Furthermore, U and R increase in two cases: (1) if $\mathbf{p}$ goes toward a line defined by $\mathbf{o}_1$ and $\mathbf{o}_2$ or (2) if $\mathbf{p}$ goes far from $\{\mathbf{o}_1, \mathbf{o}_2\}$. We also see at the bottom that our generic reliability is similar to the angle-based reliability used in [6, 35]: curves implicitly defined by constant $R(\mathbf{p})$ are very similar to circles defined by constant apical angle $(\mathbf{o}_1 - \mathbf{p}, \mathbf{o}_2 - \mathbf{p})$.

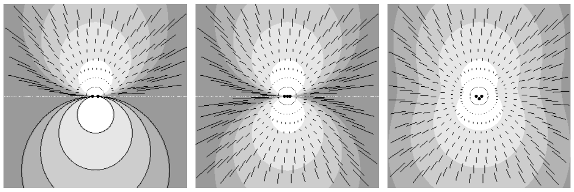

Figure 1: Virtual uncertainty U and reliability R in a plane for two ray origins (left), three collinear ray origins (middle) and three non collinear ray origins (right). Ray origins are black points in this plane. Black edges are the main axes of uncertainty ellipsoids centered at some points $\mathbf{p}$ (their length is $2U(\mathbf{p})$). Every point $\mathbf{p}$ has a gray level depending on the interval where $R(\mathbf{p})$ lies. On the left, black curves are circles defined by constant apical angles.

In the second case (in the middle), we add a ray origin $\mathbf{o}_3$ in the middle of $\mathbf{o}_1$ and $\mathbf{o}_2$. The result is unexpected: there is no improvement (i.e. U or R decrease) by adding $\mathbf{o}_3$. In fact, the results are nearly the same.

In the third case (on the right), we slightly move $\mathbf{o}_3$ toward the bottom. As expected, the improvement is noticeable in the neighborhood of the line defined by $\mathbf{o}_1$ and $\mathbf{o}_2$. In these two last cases, our R definition is naturally derived from U for any numbers of views. This is not the case for angle-based reliability [6], which is only defined for two views.

Last, a numerical test is provided for the asymptotic relation (Eq. 90) between reliability R and apical angle $(\mathbf{o}_1-\mathbf{p},\mathbf{o}_2-\mathbf{p})$ in two cases: two ray origins $\mathbf{o}_1$ and $\mathbf{o}_2$, and three ray origins with $\mathbf{o}_3=\frac{1}{2}(\mathbf{o}_1+\mathbf{o}_2)$. Function

$$a(\mathbf{p}) = (\mathbf{o}_1-\mathbf{p},\mathbf{o}_2-\mathbf{p})\frac{R(\mathbf{p})}{\sigma_\alpha\sqrt{2\mathcal{X}_2^3(p)}} \qquad (98)$$

should have a small standard deviation σ_a and a mean $\bar{a}$ close to 1 if $\mathbf{p}$ is far enough from ray origins. Values σ_a and $\bar{a}$ are estimated from samples $\mathbf{p}$ in a ring centered on $\mathbf{o}_3$ with large radius $r_l = 10||\mathbf{o}_1-\mathbf{o}_2||$ and small radius $r_s=\frac{1}{2}||\mathbf{o}_1-\mathbf{o}_2||$ (the ring is in a plane which contains the ray origins). We obtain $(\bar{a},\sigma_a)=(1.06,0.15)$ in the two ray origin case and $(\bar{a},\sigma_a)=(1.05,0.11)$ in the three ray origin case. Larger r_s and r_l improve the results, as expected.

7.2 Errors in Unit Sphere, Image and 3D

We have seen that the generic covariance definition is based on $\mathbb{S}^2$ error (Section 2). Furthermore, the projection function p of the camera propagates this error to image error such that the uncertainty ellipses of image error are distortion ellipses of p (Section 3.1). Figure 2 shows these ellipses for our equiangular catadioptric camera (on the right) and two other cases: a perspective projection into a cube face (on the left) and an equirectangular projection (in the middle). The equirectangular projection maps 3D point to its spherical coordinates (φ,θ) and is often used for panoramic imaging (Figure 2 only shows a half part of the view field). We see that the propagated image error depends on p and is not "standard" (isotropic and uniform in the whole image), especially for cameras with wide view fields.

We deduce from Figure 2 that a local rectification from the catadioptric image (on the right) into cube face (on the left) provides a more standard image error. A yet more standard image error is obtained if we replace the perspective projection (left of Figure 2) by the equirectangular projection (right of Figure 3).

Figure 3 illustrates the virtual cubes and the epipolar geometry involved in the dense stereo step of our reconstruction method (Section 5.2). The equirectangular projection is used for the local rectifications into cube faces. According to Section 3.2, this choice makes the generic covariance similar to the virtual covariance.

7.3 Accuracy of Local 3D Models for a Synthetic Scene

Now, quantitative results are given for a local 3D model reconstructed from catadioptric images: scene accuracy (discrepancy between scene reconstruction and its ground truth) and calibration accuracy (discrepancy between the calibration and its ground truth).

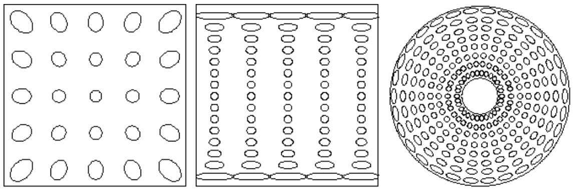

Figure 2: Distortion ellipses for the perspective projection into a face cube (left), the equirectangular projection (middle) and our equiangular catadioptric camera (right). The circular cone aperture 2ϵ of all distortion ellipses is $\frac{\pi}{25}$.

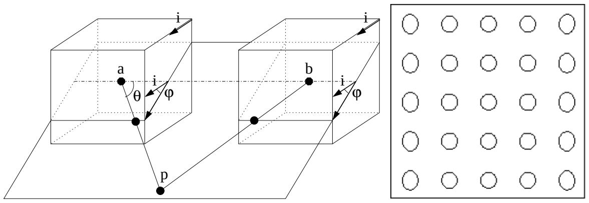

Figure 3: Left: two virtual cubes with pair-wise parallel faces for dense stereo, and epipolar plane defined by point $\mathbf{p}$ and camera centers $\mathbf{a}$ and $\mathbf{b}$. The projection of $\mathbf{p}$ into the front face of the left cube is parametrized by spherical coordinates φ and θ. Right: distortion ellipses with aperture $2\epsilon = \frac{\pi}{25}$ for the projection function $\mathbf{p} \mapsto (\theta(\mathbf{p}), \varphi(\mathbf{p}))$ into cube face.

Summary First, synthetic images are generated using ground truth (non-central) calibration. Second, structure-from-motion [18] (SfM) and the 3D modeling method described in Section 5 are applied using many central calibrations defined by radial distortion functions. Third, the scene reconstructions are registered with ground truth and quantitative evaluations are made. Details on non-central calibration and registration are shown later in Appendix A. The scene reconstructions are degraded for many reasons: image noise, approximate calibration, small baseline (due to distant or collinear points with the camera motion) and low textured areas.

Experiment Choice The catadioptric camera moves inside a textured cube to experiment the reconstruction in the whole field of view. A representative range of baselines is obtained with the following ground truth: the $[0,5]^3$ cube and camera locations defined by $\mathbf{o}_i = \begin{pmatrix}1 & 1+\frac{i}{5} & 1\end{pmatrix}^\top, \quad i \in \{0,1,2\}$ (numbers in meters). Camera orientations (rotations) are slightly perturbed around $\mathtt{I}_3$ to simulate hand-held camera. The catadioptric image is a ring with large radius of 1128 pixels and is projected into cube faces such that each face has about 440 × 440 pixels (as in real experiments).

Scene Accuracy A standard definition of error $err(\mathbf{p})$ for a reconstructed point $\mathbf{p}$ is the signed distance of $\mathbf{p}$ to the ground truth surface. In the context of a camera rotated around a small object, it is a natural choice to tolerate a uniform error for all parts of the object [31]. In our context of a camera moving in a scene including close foreground and far background, it is more adequate to define an error tolerance which increases with point depth. Thus, our accuracy measure $a_{0.9}$ is defined by the 90% fractile of $\frac{|err(\mathbf{p})|}{||\mathbf{p}-\mathbf{o}_1||}$ for all triangle vertices $\mathbf{p}$ of the 3D model. In other words, $|err(\mathbf{p})| \leq a_{0.9}||\mathbf{p} - \mathbf{o}_1||$ is true for 90% of vertices.

Experiments Figure 4 shows reconstruction results of the synthetic cube using three central calibrations "Best", "Shift" and "Sfm". These calibrations are defined by radial distortion functions $\beta(r)$, which

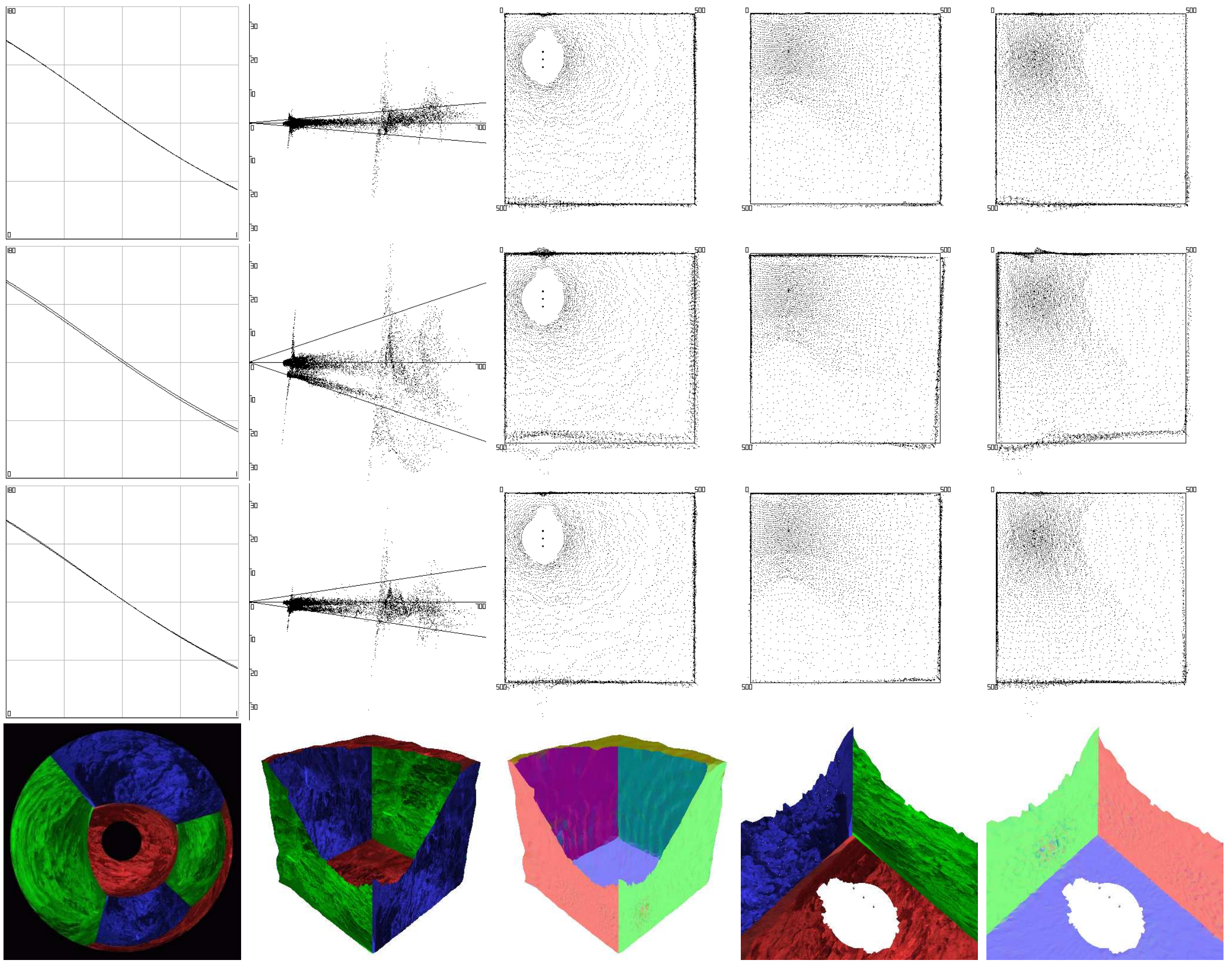

Figure 4: Reconstruction results of a synthetic cube using three calibrations: “Best” (row 1), “Shift” (row 2), and “Sfm” (row 3). Calibrations are defined by radial distortion functions which map radius r of image point to angle β between ray and camera axis. From left to right: mapping $r \mapsto \beta(r)$ superimposed with its ground truth, error graph, triangle vertices and faces of ground truth cube projected onto 3 planes (more details in the text). All numbers are in centimeters, except in the first column. Row 4 shows many texture-mapped views of the local 3D model reconstructed with “Sfm” and one image of the sequence. Each local 3D model is constructed from 3 camera locations, which are shown in the middle column.

map normalized radius r of image point to angle β between ray and camera axis (small and large circles are $r = 0$ and $r = 1$, respectively).

The first column shows many $\beta(r)$ graphs. Camera poses are estimated using Structure-from-motion [18], and local 3D models are obtained using method of Section 5. “Best” is a very accurate approximation of radial distortion defined by the true (non-central) calibration. “Shift” is “Best” shifted by 2 degrees. “Sfm” is estimated by structure-from-motion [18]. The standard deviation of β (with ground truth) of calibrations “Best”, ”Shift” and “Sfm” are 0.077, 1.98 and 0.65 degrees, respectively.

The second column of Figure 4 shows the corresponding error graphs defined by joint distribution of depth $x = ||\mathbf{p} - \mathbf{o}_1||$ and error $y = err(\mathbf{p})$ for all triangle vertices $\mathbf{p}$ of the 3D models. 90% of dots (x, y) are between the two lines $y = \pm a_{0.9}x$ with $a_{0.9}^{best} = 0.0085$, $a_{0.9}^{shift} = 0.033$ and $a_{0.9}^{sfm} = 0.015$. In examples “Shift” and “Sfm”, $\beta(r)$ inaccuracy produces a majority of vertices inside the ground truth cube where error $err(\mathbf{p})$ is negative. The $\beta(r)$ inaccuracy also increases $a_{0.9}$. The cloud dispersion of “Sfm” is slightly larger than that of “Best”.

The other columns of Figure 4 project the local 3D models into planes (xy), (xz) and (zy), which are parallels to faces of the ground truth cube. In the “Shift” case, calibration inaccuracy produces camera pose distortion, cube distortions and a majority of vertices inside the ground truth cube. These problems are less visible for “Sfm”.

The row on the bottom shows texture-mapped views and triangle orientations of the reconstructed model “Sfm”. A large neighborhood of a cube corner (on the left) and a small part of the red ground (on the right) are not reconstructed since they are not in view field.

7.4 Local 3D Model Selection for Catadioptric Camera

The global 3D model of a scene is obtained by view point selection of local 3D models (Section 6.1): a point $\mathbf{p}$ reconstructed in local model l_0 is retained in the global model if the relative uncertainty $U^r_{l_0}(\mathbf{p})$ is less than a threshold (Eq. 97). Section 7.4 illustrates how this method works for a typical list of local 3D models (one local model is reconstructed for each triplet of consecutive images of the sequence).

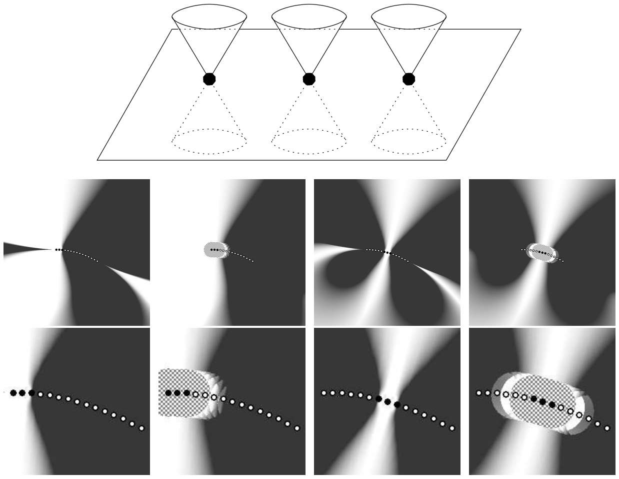

Figure 5: Top: three locations (black points) of a catadioptric camera moving in horizontal plane (view fields are bounded by cones). Middle: mapping $\mathbf{p} \mapsto U^r_{l_0}(\mathbf{p})$ on this plane and a shifted plane with l_0 two 3-view local 3D models of a 15-view sequence. Bottom: local zooms of figures in the middle. The camera locations of local model l_0 are black points; others locations are white. Values $[1, 1.5]$ of $U^r_{l_0}$ are mapped to colors white-gray; larger values are dark gray; undefined $U^r_{l_0}(\mathbf{p})$ area is white-gray checkerboard.

Figure 5 shows values of $U^r_{l_0}$ for a catadioptric camera pointing toward the sky and moving on the horizontal ground with almost collinear camera locations. In the first case (columns 3 and 4 on the right), the local model is not at the end of the whole sequence. We note that a kind of planar slice of the 3D space contains small values of $U^r_{l_0}(\mathbf{p})$, with $U^r_{l_0}(\mathbf{p}) = 1$ at the central component. The planar slice goes across the middle camera, its thickness increases with the distance to the middle camera, and it is connected to both ends of the whole camera sequence. It should be remembered that visibility is used in the view point selection as follows: we reset $U_l(\mathbf{p}) = +\infty$ if $\mathbf{p}$ is not in the view field (of at least one image) of local model l. In the catadioptric case, we have $U_l(\mathbf{p}) = +\infty$ if $\mathbf{p}$ is in a blind cone of l shown on the top of Figure 5. This has several consequences: (1) $U^r_{l_0}$ is not defined outside the view field of l_0, (2) $U^r_{l_0}$ is not continuous at boundaries of view fields of l with $l \neq l_0$, and (3) the reconstruction of ground surface is allowed. Figure 5 shows these 3 consequences in column 4: (1) white-gray checker-board, (2) gray level discontinuities and (3) low values of $U^r_{l_0}$ in the immediate neighborhood of the camera trajectory on the ground (if the shifted plane of the figure is the ground surface). In the second case (columns 1 and 2 on the left), the local model is at the end of the whole sequence. The planar slice is replaced by a large section of a half space, and the three notes on visibility are still correct.

Figure 6 shows results of view point selection (Eq. 97 with $\epsilon = 0.1$) combined with unreliable point rejection (Eq. 94 with $R_{max} = 0.1$, $\sigma_\alpha = 0.001$ and $\mathcal{X}^2_3(0.9) = 6.25$). On the left, gray levels encode the minimal uncertainty available to reconstruct scene point $\mathbf{p}$ with all local 3D models of the typical list L of a 11-view sequence. On the right, gray levels encode the number of local models available to reconstruct scene point $\mathbf{p}$ thanks to view point selection. This number is the modeling redundancy at $\mathbf{p}$. We note that redundancy increases if $\mathbf{p}$ leaves the camera trajectory, and redundancy is limited thanks to unreliable point rejection.

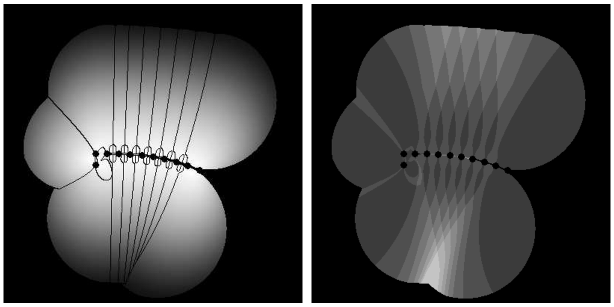

Figure 6: Left: gray levels encode $\mathbf{p} \mapsto \min_{l\in L} U_l(\mathbf{p}) = U_{l(\mathbf{p})}(\mathbf{p})$ in the plane where the camera moves. Areas with the same best local model $l(\mathbf{p})$ are bordered by black lines. Right: gray levels encode the number of local models accepted by view point selection (darkest gray: 1 local model, white: 8 local models). Black pixels on the image borders are points rejected by reliability condition.

7.5 Real Examples

Now, the overall reconstruction system is experimented on still images sequences taken by equiangular catadioptric cameras, which are hand-held and mounted on a monopod (such a camera choice is discussed in Section 1.3). We obtain 3264 × 2448 JPEG images with the following setup: adequate mirrors [1] mounted with the Nikon Coolpix 8700 using adapter rings (Figure 7). Both image dimensions are divided by 2 to accelerate structure-from-motion (SfM), mesh estimations, and to facilitate texture storage of the VRML models. Only dense stereo benefits from original image sizes using the local rectifications into cube faces (Section 5.2).

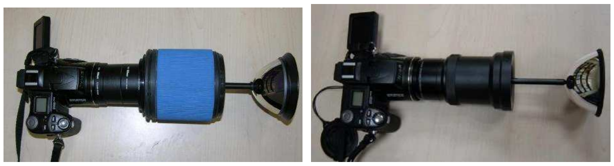

Figure 7: 360 One VR (left) and 0-360 (right) mirrors mounted with the Nikon Coolpix 8700.

Church Sequence We take 208 images during a complete walk around a church using the 360 One VR (model 3) mirror, using a supplementary adapter ring which adds 2 cm between mirror and perspective camera. The trajectory length is about $(25 \pm 5cm) \times 208 = 52 \pm 10m$ (the exact step lengths between consecutive images are unknown). The radii of large and small circles of the catadioptric images are 1128 and 232 pixels.

SfM is the first step of the method (Section 1.4). Figure 8 shows top views of the SfM results obtained with slight modifications of the method [18]: the number of inlier points is bounded by 500 in each image to accelerate hierarchical bundle adjustment and (optional) loop closure, then a last global bundle adjustment is applied without inlier bounds. The unclosed reconstruction and its drift are shown in the top left corner; its closed version is shown in the top-right corner. The final result is shown at the bottom with some images. It has 76033 3D points and 477744 points in images satisfying the geometry. The final RMS error is 0.74 pixels. A 2D point is considered as an outlier if its reprojection error is greater than 2 pixels. With our setup, the estimated view field angles are $\alpha_{up} = 41.5, \alpha_{down} = 141.7$ degrees (the angles given by the mirror manufacturer are $\alpha_{up} = 40, \alpha_{down} = 140$ degrees).

The second step is the calculation of a list of local 3D models (Section 5). Here we use typical list: one local model for each triple of consecutive images of the sequence. Figure 9 shows the reference image, the depth map and the 2D mesh of a local model. The reference image is between the two others. Once the catadioptric images are projected into cube faces, quasi-dense propagation [20] is used and benefits by the large number of seed matches provided by SfM inliers. The depth map is given in the original reference image. The 2D mesh has 22894 triangles and the 2.5D mesh is generated using $\mathcal{X}_3^2(0.9) = 6.25$ and $\sigma_\alpha = 0.00082$ radians. This value of σ_α is estimated from the dense cloud reconstruction (using Eq. 60) of all local models of the list. Figure 9 also shows views of the local model without and with rejection of unreliable triangles (using Eq. 94 and $R_{max} = 0.05$). The local model with rejection has 10510 triangles. A small part of the ground and the upper part of the facade are in the blind cones defined by the small and large circles of the catadioptric images and can not be reconstructed for this reason.

Once the 208 local 3D models of the closed sequence are reconstructed (with the same value of σ_α), the global 3D model is obtained by view point selection and redundancy reduction (Section 6). First, triangles of local models are selected using the reliability test: all vertices $\mathbf{p}$ of a triangle should satisfy Eq. 94 with $R_{max} = 0.04$. We obtain 2249675 triangles. Second, view point selection is applied: any triangle of any local model is retained in the global model if it has at least one vertex $\mathbf{p}$ such that Eq. 97 is true with $\epsilon = 0.1$. Only 31% of triangles are retained (681561 triangles). The (final) global model has 368814 triangles after redundancy reduction. Figure 10 shows many views of this global model of the church and its neighborhood. The reader can match the top view of the model (top of Figure 10) with the top view of the SfM result in Figure 8.

Now, the difficulties of this scene and the consequences on the global 3D model are shown in detail. There are four trees in the immediate neighborhood of the church and many others which are more distant. Since trees are important scene components, we choose to not remove triangles which are unconnected with others to modelize the foliage as clouds of textured triangles. Thus, the “Triangle Damping” mesh operation is preferred and used instead of the “Triangle Removal” operation (Section 5.5). The drawback is the presence of certain isolated and false triangles in other scene components. Furthermore, the scene has many low textured areas: streets, cars, and sky. The consequences are the presence of a number of holes and inaccurate reconstructions in the streets or distant cars. Reconstruction is almost impossible for the blue car which is very close to the camera trajectory (see picture in the top-left corner of Figure 8). The sky is globally unmatched thanks to the confidence measure used by quasi-dense propagation [20], but minor parts of the sky occur in reconstructed foliage and at the top of some buildings. Last, the images are taken in the early morning. Two consequences are: (1) the shadow of the author occurs in several images and (2) the first and last images of the closed sequence have different textures. The former

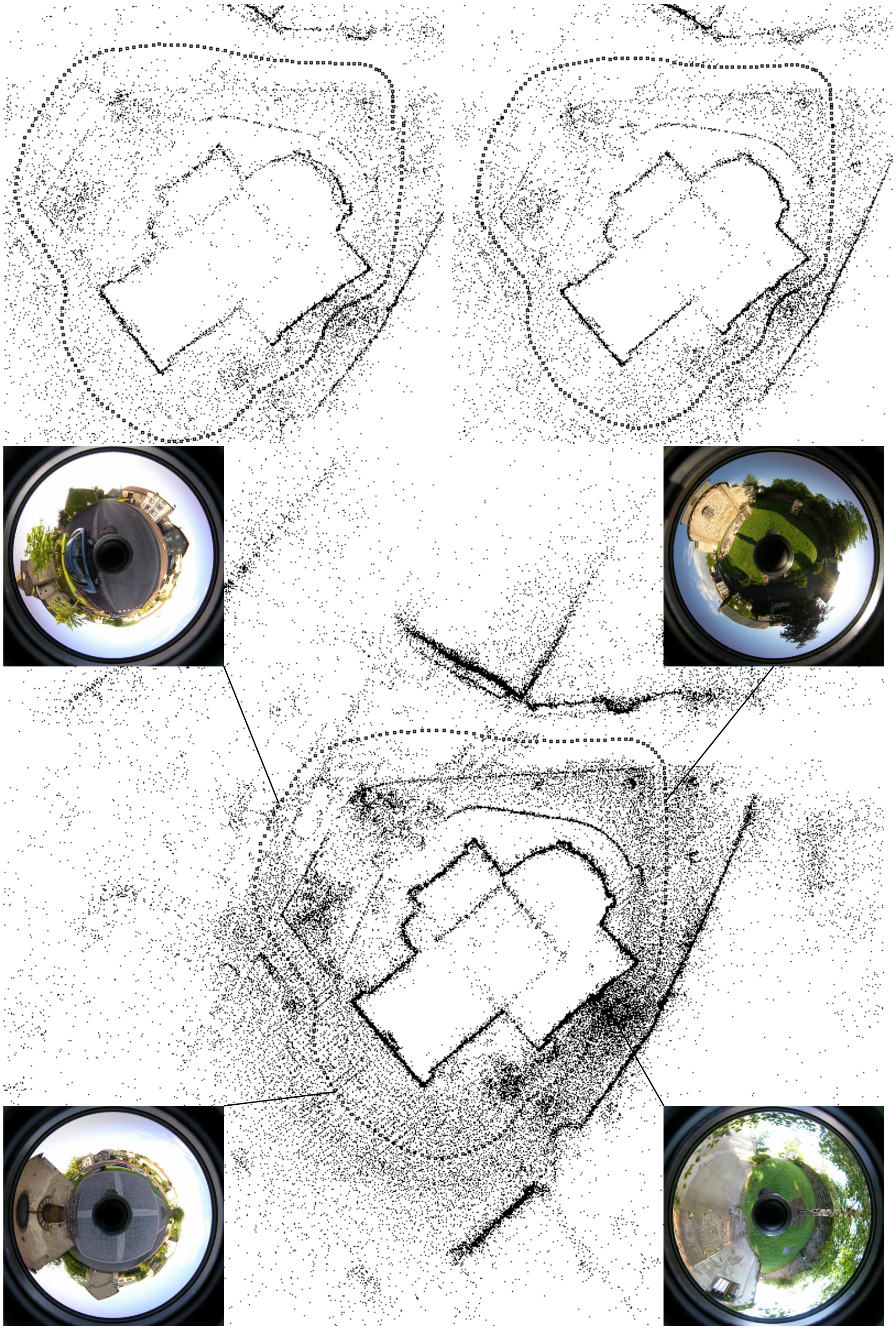

Figure 8: Top: unclosed and closed structure-from-motion results with reduced number of 3D points. Bottom: closed structure-from-motion with all 3D points and some images of the Church.

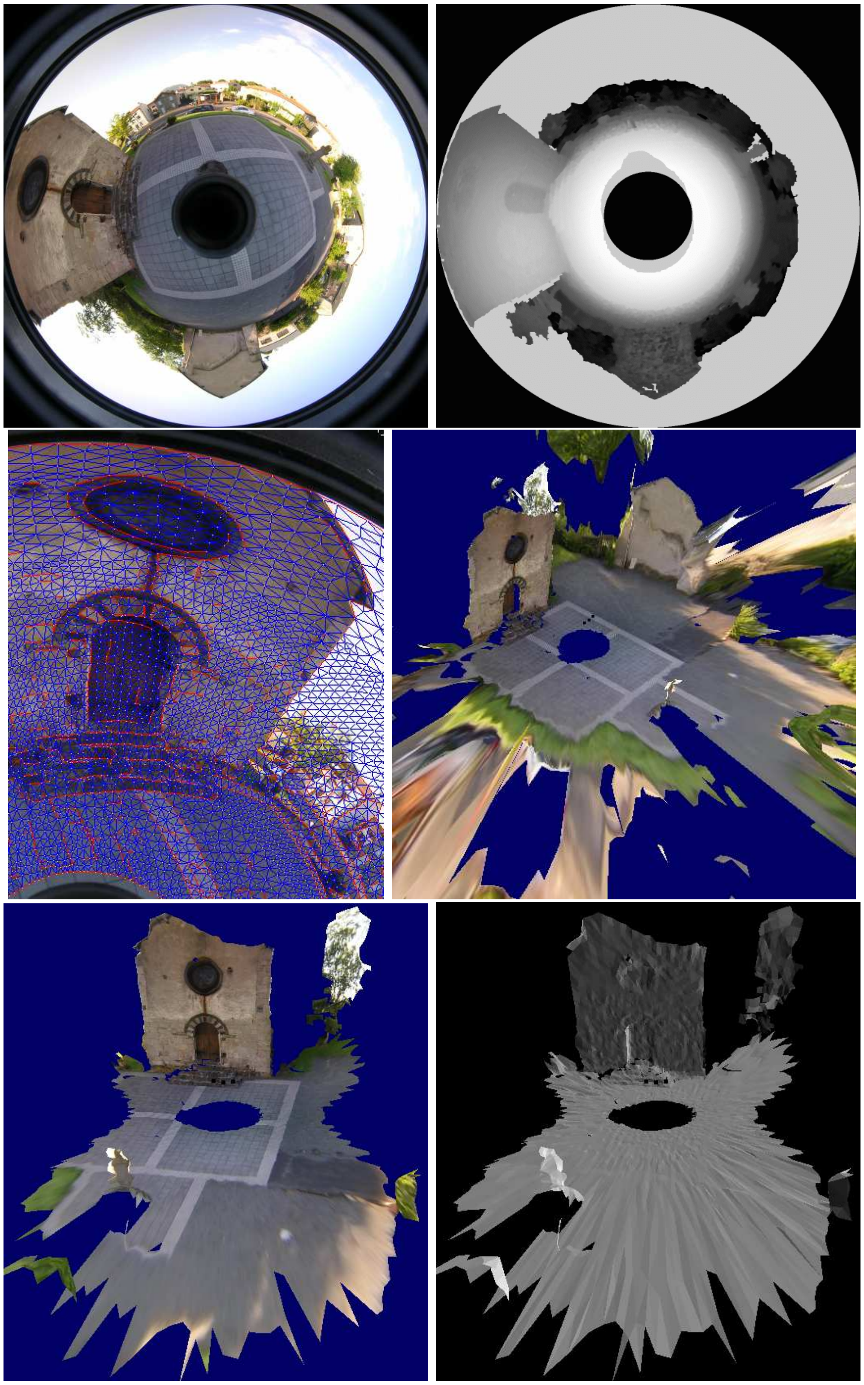

Figure 9: Top: reference image and depth map of a 3-view local model. Middle left: local view of the 2D mesh (constrained and unconstrained Delaunay edges are red and blue, respectively). Middle right: local model without reliability test. Bottom: local model with reliability test (triangle orientations on the right).

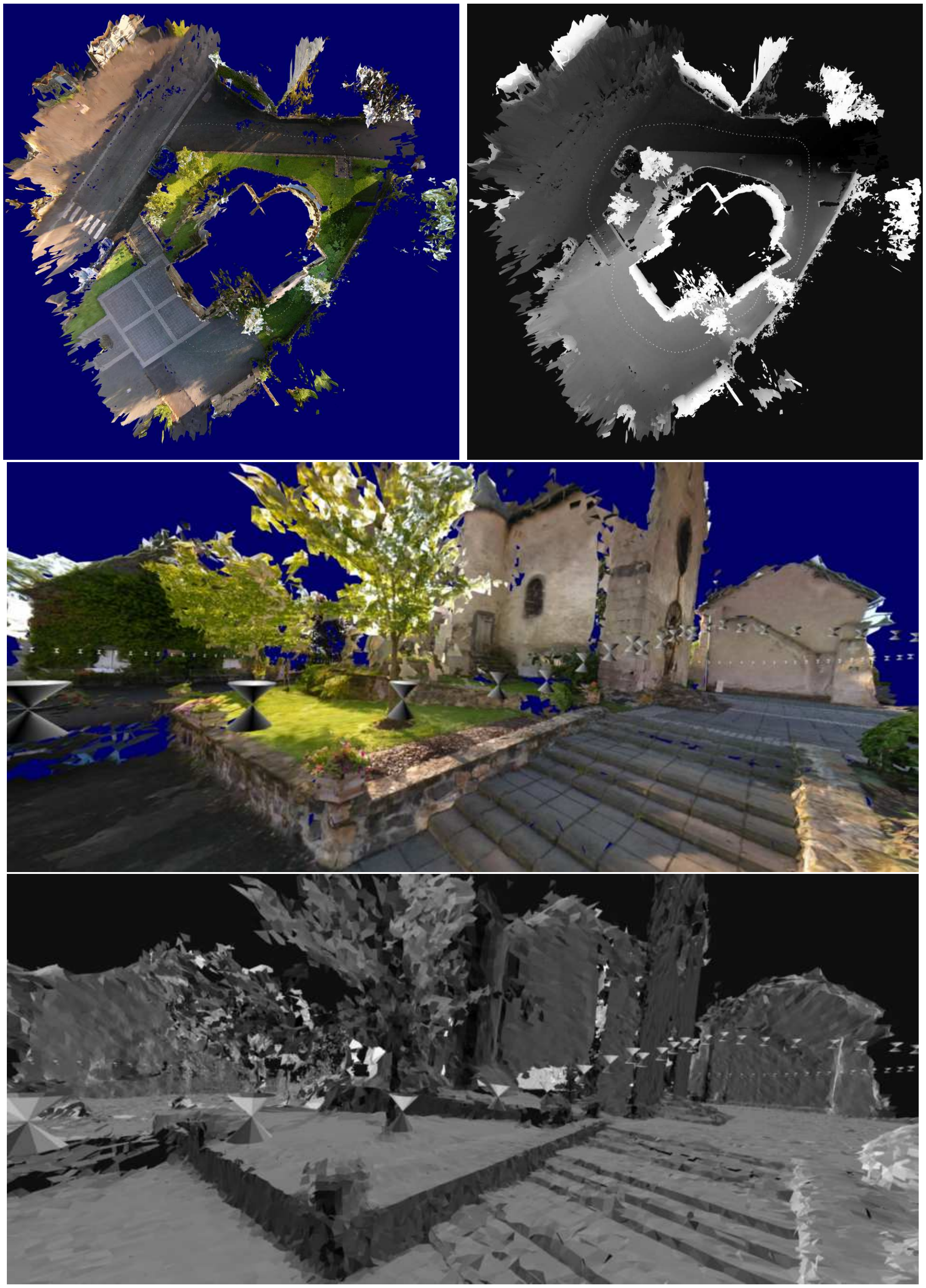

Figure 10: From top to bottom: top view and height map of church global model, local view (texture and orientation of triangles).

complicates dense stereo and ground reconstruction. The latter complicates matching used by SfM loop closing and dense stereo.

Old Town Sequence This sequence has 354 images and is acquired using the 0-360 mirror. The radii of large and small circles are 1140 and 204 pixels. The trajectory length is about $(35 \pm 5cm) \times 353 = 122 \pm 17m$.

SfM estimates 149792 3D points with 819087 2D inliers and RMS error equal to 0.75 pixels. These inlier numbers are larger than those in [17] thanks to a slight modification of the SfM method. Figure 11 shows a top view of the SfM result and some images of the sequence. The estimated view field angles are $\alpha_{up} = 34.5, \alpha_{down} = 152.9$ degrees (the angles given by the mirror manufacturer are $\alpha_{up} = 37.5, \alpha_{down} = 152.5$). Then, the 352 local models of the typical list are estimated using $\mathcal{X}_3^2(0.9) = 6.25$ and $\sigma_\alpha = 0.00089$ radians. The numbers of triangles after unreliable triangle rejection ($R_{max} = 0.03$), view point selection ($\epsilon = 0.1$) and redundancy reduction are 4991183, 1786919 and 1184620, respectively. Figure 12 shows many views of the final global model. The joint video is a scene walkthrough in the scene.

We note that the reconstruction of building tops is inaccurate since it involves several difficulties. First, building tops are not visible by the closest local model but by others which are more distant (small baseline). Second, there is a lack of calibration accuracy (angle α_{up}) in the neighborhood of the large circle where building tops are projected . Third, false matches are not rejected by epipolar constraint since the camera motion is parallel to the building borders. In this context, unreliable triangle rejection is useful to avoid these parts in the global model. Other difficulties are low textured areas (buildings, ground, cars), specular reflections on windows, and a major illumination change during image acquisition.

We also note that a few parts of the scene are not retained in the global model by view point selection, although they are not in a blind cone of the local model which provides the smallest uncertainty. These parts sometimes include walls with plane normal parallel to the camera trajectory at the closest camera location. This is due to our current and limited use of visibility in the view point selection (end of Section 6.1).

8 Conclusion

This paper is a new step towards the automatic 3D modeling of scenes using a generic method, which can be applied for any kind of camera. First, we introduce a generic covariance for central cameras. Proof is given for its definition and several properties: asymptotic behaviors, links with the apical angles, links between the generic covariance and the covariance which results from the ray intersection problem defined by the sum of squared reprojection errors in pixels. Second, we present a 3D modeling method from images which systematically uses the generic covariance. More precisely, generic covariance is used to weight minimized scores involved in triangle estimation, to decide connections between triangles, to fill holes, to reject unreliable triangles, to define view point selection and to reduce model redundancy.

However, these contributions do not go far enough for non-central camera. The calculation of generic covariance at point $\mathbf{p}$ requires the ray origins corresponding to $\mathbf{p}$. These ray origins are known if the camera is (approximated by) a central camera, but a generic method is still needed to obtain ray origins for a non-central camera. Other limitations are due to the use of 2D meshes in images: the camera should not be too exotic to back-project connected image points (2D triangles) to connected points of the scene surface. Furthermore, the current implementation of 2D mesh initialization is only done for our cameras. All these limitations are topics for future research.

Experiments are another contribution of the paper. Our context is useful for applications and has rarely been addressed until now: the automatic 3D modeling of scenes using catadioptric cameras. Once structure-from-motion has estimated camera calibration and poses from image sequence, the 3D modeling method based on generic covariance is applied to generate a global 3D model of the scene. Experiments illustrate generic covariance properties. They include accuracy estimation of our catadioptric setup on syn-

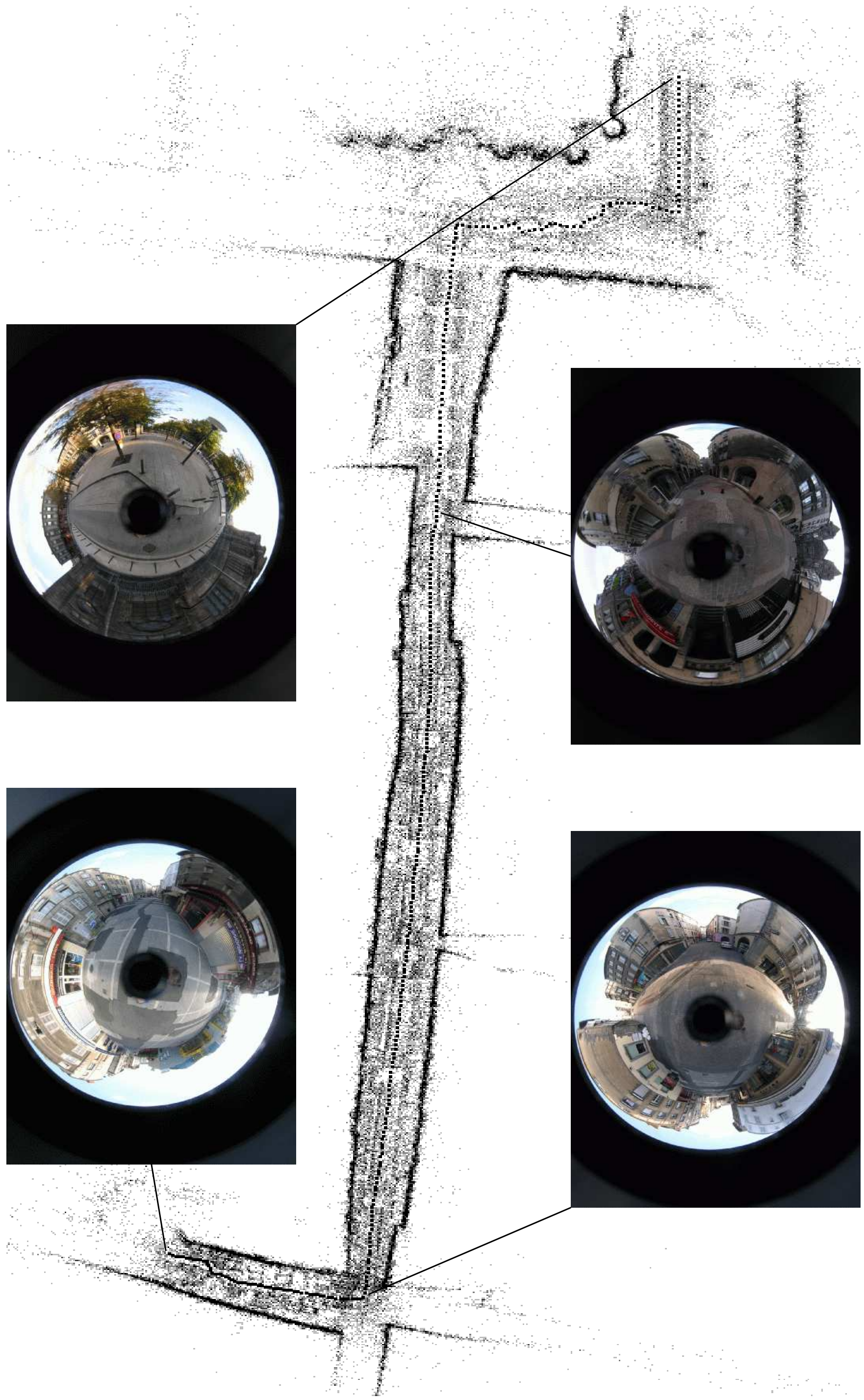

Figure 11: Top view of the structure-from-motion result with some images of the old town.

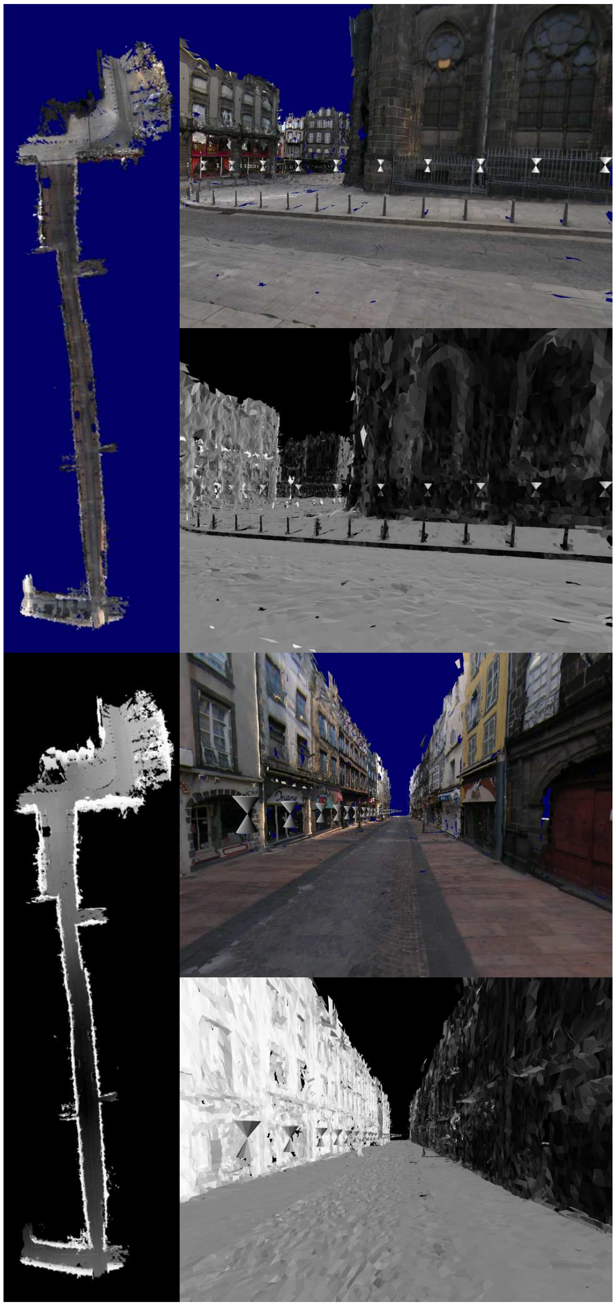

Figure 12: Left: top view and height map of the old town global model. Right: local views (texture and orientation of triangles).

thetic images, and provide 3D model reconstructions from real sequences with hundreds of images.

Many improvements are possible: a better use of visibility in view point selection, applying merging methods like [4, 32] after view point selection. It is important to remember that view point selection is a key issue for 3D modeling, but it can not replace (and it is not a competing method of) such merging methods. Lists of local models with different baselines would also be useful to increase accuracy. Last, several steps of the reconstruction method may be improved (image matching, texture merging) and other steps may be added (mesh simplification, image gain corrections).

Appendix A: Data for Synthetic Experiments

This appendix explains how to obtain a realistic non-central calibration of a catadioptric camera. This is useful for the synthetic experiments in Section 7.3. Here we use the 0-360 mirror [1] mounted with the Nikon Coolpix 8700. Non-central calibration is required to project the synthetic scene in images by taking into account ray reflexion on the mirror. It is defined by the mirror profile and the matrix of the perspective camera in the coordinate system of the mirror [18].

First, the mirror profile has to be estimated since it is not provided by the mirror manufacturer. Profile $z(r)$ is used by the mirror cylindrical parametrization

$$f(r,\theta) = \begin{pmatrix} r\cos(\theta) & r\sin(\theta) & z(r) \end{pmatrix}^\top \qquad (99)$$

and is obtained as follows: (1) select a room with lighting reduced to a single bulb on the ceiling (2) put a Cartesian graph paper on the ground such that the light rays are orthogonal to the paper (3) put the mirror on the paper such that the mirror symmetry axis is horizontal and parallel to one of the main directions of the Cartesian graph paper (4) mark several points on the paper at the border of the mirror shadow and (5) fit polynomial $z(r)$ from these points. We obtain

$$z(r) \quad = \quad 0.0186 + 0.06188r + 0.10812r^2$$

$$+ \quad 0.0312154r^3 \text{ with } 0 \leq r \leq 3.9 \text{ cm} \tag{100}$$

Second, the matrix of the perspective camera is estimated. A value of the focal length f_p is available from the EXIF data in the JPEG image returned by the perspective camera: $f_p = 7094$ pixels. Furthermore, a typical radius of the large circle in catadioptric images is $r_{up} = 1128$ pixels. Following Section 4.5 of [18], we obtain the z-coordinate of the perspective camera center on the mirror symmetry axis using the Thales relation $\frac{3.9}{z(3.9)+z_p} = \frac{r_{up}}{f_p}$. Now, we obtain the perspective camera matrix $\mathtt{K}_p\mathtt{R}_p^\top \begin{pmatrix} I_3 & -\mathbf{t}_p \end{pmatrix}$ in the mirror coordinate system: intrinsic parameters $\mathtt{K}_p = \text{diag}(f_p, f_p, 1)$, rotation $\mathtt{R}_p = \mathtt{I}_{3\times3}$ and perspective center $\mathbf{t}_p = \begin{pmatrix} 0 & 0 & -z_p \end{pmatrix}^\top$ with $z_p = 20.8$ cm.

Third, we check the view field defined by these values. We obtain $\alpha_{up} = 37.4$ using ray reflection at the large circle. In practice, the ratio between radii of small and large circles in the catadioptric image is equal to 0.18. This ratio and ray reflection at the small circle imply $\alpha_{down} = 153.0$. These angle values are similar to those of the mirror manufacturer ($\alpha_{up} = 37.5$ and $\alpha_{down} = 152.5$ degrees).

We also define the 3D registration which maps the estimated scene reconstruction to scene ground truth. This is useful since the definition of reconstruction accuracy in Section 7.3 depends on it. This registration is a similarity transformation defined as follows. Its Euclidean part maps the local basis of the 2^{nd} pose of the reconstruction (central camera model) to the local basis of the mirror at the 2^{nd} pose of the ground truth (non-central camera model). Its scale part is the ratio of trajectory lengths between ground truth and reconstruction bases.

Appendix B: 2D Mesh

This appendix is a summary of the 2D mesh method used in Section 5.1. There are three steps.

Mesh Initialization First, a Delaunay triangulation is initialized in the reference image such that the solid angles of all triangles are roughly the same. This step is defined as follows for our catadioptric cameras (the general case has been left out for future work). We generate a checkerboard such that (1) each cell is quadrilateral and has two triangles (2) the mesh resolution is defined by a mean length of cell edges equal to 8 pixels (3) catadioptric image borders enforce constrained edges on the Delaunay and enforce the global shape of the checkerboard. At this point, cell rows of the checkerboard are concentric rings and cell columns are radial sections. A modification of the current mesh is useful to increase the uniformity of solid angles of triangles: we remove one vertex out of two at image border neighborhoods to increase the solid angles of triangles in these areas.

Gradient Edge Integration Second, the gradient edges are integrated in the mesh by moving mesh vertices slightly and forcing mesh edges to be constrained. We have not taken into account all gradient edges since the mesh resolution has been previously fixed. So gradient edges are integrated in a best first order. A contour is a list of connected pixels which have maximum local image gradient. The contour score is equal to the sum of gradient modulus for all its pixels. We pick the contour with the highest score, and find the list of closest vertices to its pixels such that these vertices have not been used before for any other contour. Then two consecutive vertices are moved slightly to approximate the contour if the part of the contour between vertex ends is a segment. Once all contours have been considered by decreasing score, a completion step is used in order to try to constrain new mesh edges if they approximate a contour in their immediate neighborhood.

Mesh Refinement Third, the 2D mesh is refined by alternating "continuous" operations (move vertices to minimize a global cost combining color variance in triangles and mesh smoothness) and "discrete" operations (flip edges and merge vertices to improve aspect ratio of triangles).

The continuous operation is useful for many reasons. First, a few gradient edges may be missed by the previous step, and minimizing the sum of color variances for each triangle is an other way to increase the probability that the gradient edges are on the mesh edges. Second, the gradient edge integration deformed the initial mesh only locally such that the constraint of a same solid angle for all triangles is highly violated. Minimizing the mesh smoothness (sum of squared modulus of an umbrella operator) is a way to incite incident triangles to have similar solid angles. Minimizing the mesh smoothness is also useful to improve triangle aspect ratio and regularize the minimization of color variance.

The cost function $e_{2d}(\{p_v\})$ is defined by

$$\sum_{p \in t \in T} ||c_p - \sum_{p' \in t} \frac{c_{p'}}{|t|}||^2 + \lambda \sum_{v \in V} || \sum_{v' \in \mathcal{N}_v} p_v - p_{v'}||^2 \quad (101)$$

and T the list of mesh triangles, $|t|$ the area of triangle t, V the list of mesh vertices, $\mathcal{N}_v$ the list of vertices which are connected to v by a mesh edge. Color c_p at pixel p is RGB, p_v is the image location of vertex v and λ is equal to 1000. The vertex locations $\{p_v\}$ are the parametrization of e_{2d}, and e_{2d} is minimized using a descent method. All mesh vertices are allowed to move in 2D, except vertices which are incident to a constrained edge (vertices at gradient edges). The latter are only allowed to move in 1D along the detected gradient edges. This gives priority to detected gradient edges over the minimization of color variance, which may sometimes be contradictory.

References

[1] www.0-360.com, www.kaidan.com.

[2] A. Akbarzadeh, J. Frahm, P. Mordohai, B. Clipp, C. Engels, D. Gallup, P. Merell, M. Phelps, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch, H. Towles, D. Nister, and M. Pollefeys. Towards urban 3d reconstruction from video. In *3DPTV'06*, 2006.

[3] R. Bunschoten and B. Krose. Robust scene reconstruction from an omnidirectional vision system. *IEEE Transactions on Robotics and Automation*, pages 351–357, 2003.

[4] B. Curless and M. Levoy. A volumetric method for building complex models from range images. *SIGGRAPH*, 30, 1996.

[5] K. Daniilidis. The page of omnidirectional vision. www.cis.upenn.edu/˜ kostas/omni.html.

[6] P. Doubek and T. Svoboda. Reliable 3d reconstruction from a few catadioptric images. In *OMNIVIS'02*, 2002.

[7] J. Evers-Senne, J. Woetzel, and R. Koch. Modelling and rendering of complex scenes with a multi-camera rig. In *CVMP'04*, 2004.

[8] O. Faugeras, Q. Long, and T. Papadopoulos. *Geometry of Multiple Images*. MIT Press, 2001.

[9] S. Fleck, F. Busch, P. Biber, W. Strasser, and H. Andreasson. Omnidirectional 3d modeling on a mobile robot using graph cuts. In *ICRA'05*, 2005.

[10] M. Grossberg and S. Nayar. A general imaging model and a method for finding its parameters. In *ICCV'01*, 2001.

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

[12] D. F. J. Barron and S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1), 1992.

[13] S. Kang and R. Szeliski. 3-d scene data recovery using omnidirectional multibaseline stereo. *IJCV*, 25(2), 1997.

[14] J. Kannala and S. Brandt. A generic camera model and calibration method for conventional, wide-angle and fish-eye lenses. *IEEE TPAMI*, 28(8), 2006.

[15] V. Kolmogorov and R. Zabih. Computing visual correspondances with occlusions using graphcuts. In *ICCV'01*, 2001.

[16] M. Lhuillier. Effective and generic structure from motion using angular error. In *ICPR'06*, 2006.

[17] M. Lhuillier. Toward flexible 3d modeling using a catadioptric camera. In *CVPR'07*, 2007.

[18] M. Lhuillier. Automatic scene structure and camera motion using a catadioptric system. *CVIU*, 109(2), 2008.

[19] M. Lhuillier. Toward automatic 3d modeling of scenes using a generic camera model. In *CVPR'08*, 2008.

[20] M. Lhuillier and L. Quan. Match propagation for image-based modeling and rendering. *IEEE TPAMI*, 24(8), 2002.

[21] H. Li, R. Hartley, and J. Kim. Linear approach to motion estimation using generalized camera model. In *CVPR'08*, 2008.

[22] B. Micusik and T. Pajdla. Structure from motion with wide circular field of view cameras. *IEEE TPAMI*, 28(7), 2006.

[23] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd. Generic and real time structure from motion using local bundle adjustment. *IVC*, 27, 2009.

[24] D. Nister and H. Stewenius. A minimal solution to the generalized 3-point pose problem. *JMIV*, 27(1), 2007.

[25] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE TPAMI*, 15(4), 1993.

[26] R. Pless. Using many cameras as one. In *CVPR'03*, 2003.

[27] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch. Visual modeling with a hand-held camera. *IJCV*, 59(3), 2004.

[28] S. L. S. Ramalingam and P. Sturm. A generic structure-from-motion framework. *CVIU*, 103(3), 2006.

[29] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47(2), 2002.

[30] K. Schindler and H. Bischof. On robust regression in photogrammetric point clouds. In *DAGM'03 (also LNCS 2781)*, 2003.

[31] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR'06*, 2006.

[32] M. Soucy and D. Laurendeau. A general surface approach to the integration of a set of range views. *IEEE TPAMI*, 17(4), 1995.

[33] D. Strelow, J. Mischler, S. Singh, and H. Herman. Extending shape-from-motion estimation to noncentral omnidirectional camera. In *IROS'01*, 2001.

[34] N. Tissot. Tissot's indicatrix. Wikipedia.

[35] A. Torii, M. Havlena, T. Pajdla, and B. Leibe. Measuring camera translation by the dominant apical angle. In *CVPR'08*, 2008.