AN XML FILE FORMAT FOR GENERALIZED PINHOLE CAMERAS, VERSION 1.0

Aless Lasaruk

Aless Lasaruk*
Scientific Assistant, FORWISS
Room 217, ITZ, Innstrasse 43
Universität Passau
94030 Passau (Germany)

KEY WORDS: XML, pinhole camera, computer vision, camera calibration

ABSTRACT:

This document reviews a generalized pinhole camera model with radial and tangential distortions used in many computer vision and camera calibration applications and libraries. For this model, an XML file format is specified.

1 INTRODUCTION

Modeling the behavior of a camera is a fundamental subproblem of many computer vision and multiple view geometry tasks (Hartley and Zisserman, 2000). One of the most common camera models is the generalized pinhole camera model with radial and tangential distortions. For scientific results to be reproducible, it is important to have a generally agreed format for exchanging calibration data between different libraries or scientific groups. However, such format is not established yet.

This document reviews a generalized pinhole camera model with radial and tangential distortions, as it is for example used by the OpenCV library. We specify a simple XML file format to store calibration information of a set of generalized pinhole cameras. Our focus is to present a format specification together with a precise formal description of the corresponding camera model in one document.

The outline of the document is as follows. In Section 2 the generalized pinhole camera model with radial and tangential distortions is reviewed from its mathematical point of view. Section 3 specifies the structure of an XML document to store the camera model. Section 4 summarizes our results.

2 GENERALIZED PINHOLE CAMERA

A pinhole camera essentially implements a perspective projection. A common approach to formalize such projections are homogeneous coordinates. However, since our model involves distortions, we follow an Euclidean approach in notation.

Generally, a camera $\mathcal{K}$ is a map $\mathcal{K} : S \to \mathbb{R}^2$ from a subset of *world points* $S \subseteq \mathbb{R}^3$ to the *image plane* $\mathbb{R}^2$. The properties of $\mathcal{K}$ are driven by a finite family of *calibration parameters*. A generalized pinhole camera with distortions splits as a composition of maps

$$\mathcal{K}(\mathbf{s}) = p \circ \sigma \circ \pi \circ m(\mathbf{s}). \tag{1}$$

The component maps p, σ, π, and m and the corresponding parameters are discussed below in detail.

The *camera pose* $m : \mathbb{R}^3 \to \mathbb{R}^3$ is an isometry of the three-dimensional space. More precisely, there is an orthogonal $\mathbf{R} \in \mathbb{R}^{3\times 3}$ and a $\mathbf{t} \in \mathbb{R}^3$ such that

$$m(\mathbf{s}) = \mathbf{R}(\mathbf{s} - \mathbf{t}). \tag{2}$$

The camera pose transforms points in the world coordinate system into the local coordinate system of the camera. It is easy to see that $m(\mathbf{t}) = \mathbf{0}$, in other words, $\mathbf{t}$ is the origin of the local coordinate system of the camera.

The *central projection* $\pi : \{(x, y, z) \in \mathbb{R}^3 \mid z \neq 0\} \to \mathbb{R}^2$ is defined as

$$\pi(x, y, z) = \frac{1}{z}(x, y). \tag{3}$$

This map implements the perspective projection property of the camera. It is easy to see that points $\mathbf{s} \in \mathbb{R}^3$ with $\mathbf{s} - \mathbf{t}$ orthogonal to the third row $\mathbf{r}_3$ of $\mathbf{R}$ are outside the domain of π. As all component functions except for π are defined on the complete $\mathbb{R}^2$ or $\mathbb{R}^3$ respectively, latter specifies the domain S of $\mathcal{K}$ as

$$S = \{\mathbf{s} \in \mathbb{R}^3 \mid \mathbf{r}_3^t(\mathbf{s} - \mathbf{t}) \neq 0\}. \tag{4}$$

Latter definition reveals the complementary set to S as a plane through the camera center paralel to the image plane. Typically, points there can not be observed by a real camera.

The *distortion* σ is a non-linear deformation of the plane $\sigma : \mathbb{R}^2 \to \mathbb{R}^2$. With radial and tangential distortions we have

$$\sigma(\mathbf{p}) = r(\mathbf{p}) + t(\mathbf{p}). \tag{5}$$

The radial distortions $r : \mathbb{R}^2 \to R^2$ are specified by

$$r(\mathbf{p}) = \left(1 + \sum_{i=1}^{n} c_i \|\mathbf{p}\|^{2i}\right) \mathbf{p}, \tag{6}$$

where the *number of distortion coefficients* n is typically limited 2 or 3. We follow this restriction as well. The tangential distortions $t : \mathbb{R}^2 \to \mathbb{R}^2$ are commonly limited to two parameters. In this case

$$t(x, y) = c_1 \begin{pmatrix} 2xy \\ x^2 + 3y^2 \end{pmatrix} + c_2 \begin{pmatrix} 3x^2 + y^2 \\ 2xy \end{pmatrix}. \tag{7}$$

Finally, the *pixel transform* p is a linear deformation of the plane $p : \mathbb{R}^2 \to \mathbb{R}^2$ with

$$p(\mathbf{p}) = \begin{pmatrix} \alpha & \gamma \\ 0 & \beta \end{pmatrix} \mathbf{p} + \mathbf{c}. \tag{8}$$

The entries $0 < \alpha, \beta \in \mathbb{R}$ are called the *focal parameters* of the camera. The parameter $\gamma \in \mathbb{R}$ is called the *skew* of the camera. Latter value is typically near zero. The point $\mathbf{c} \in \mathbb{R}^2$ is called the *principal point* of the camera. It is the image of the optical axis of the camera and is typically near the half of the image size in

*Corresponding author.

use.

We have just specified all required camera parameters. The parameters α, β, γ, and $\mathbf{c}$ are called the *intrinsic camera parameters* as opposed to the *extrinsic camera parameters* $\mathbf{R}$ and $\mathtt{t}$. The parameters of σ are called the *camera distortion parameters*.

3 XML SPECIFICATION

We propose to store the parameters of the camera, as defined in Section 2, directly as they appear in the camera model. Figure 1 shows a diagram of a minimal proposed XML structure for a single camera. Obviously, further attributes are possible to extend the specification.

To store multiple cameras, as for example required by a multiple view setup, an outer-most tag ⟨`calibration`⟩ is used. Figure 2 shows a diagram of a multiple view XML camera setup.

Each ⟨`camera`⟩ inside a calibration setup is given a unique identifier, which is stored in the "name"-attribute of the ⟨`camera`⟩-tag. The attribute "name" is mandatory, when multiple cameras are specified, otherwise it is optional.

For each camera the resolution of the image in use is specified. Notice that the image resolution is not part of the specification of a pinhole camera, as introduced in Section 2. However, in a multi-scale setup there may appear the need to adjust the camera to a particular scaled image size. Therefore, the resolution is stored in the ⟨`resolution`⟩-tag. The entries are assumed to have integer values.

The ⟨`projection`⟩-tag specifies the intrinsic parameters of the camera as given by Equation (8) and Equations (6) and (7). The ⟨`principle`⟩-tag defines the principal point $\mathbf{c} = (x, y)$.

The pose of the camera is specified by the ⟨`pose`⟩-tag. As with camera names, the pose is given a unique name encoded by the "name"-attribute. Including multiple poses is a common means to reduce storage when working with a moving camera with fixed intrinsic parameters. The attribute "name" is mandatory, when multiple poses are specified, otherwise it is optional.

The ⟨`rotation`⟩-tag specifies the matrix $\mathbf{R}$ in the row-major order. More precisely,

$$\mathbf{R} = \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix}. \tag{9}$$

Here, the entries are denoted exactly as they are stored in the XML structure. The ⟨`translation`⟩-tag contains the camera center $\mathtt{t}$.

4 CONCLUSIONS

We have introduced a simple but extendible XML format for specification of generalized pinhole cameras with distortions. The specification straight-forwardly generalizes to a larger number of distortions coefficients. The specification is suitable to represent multiple view setups with different cameras as easily as motion paths of a single moving camera. Finally, the specification is ready for use in a multi-scale framework.

REFERENCES

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

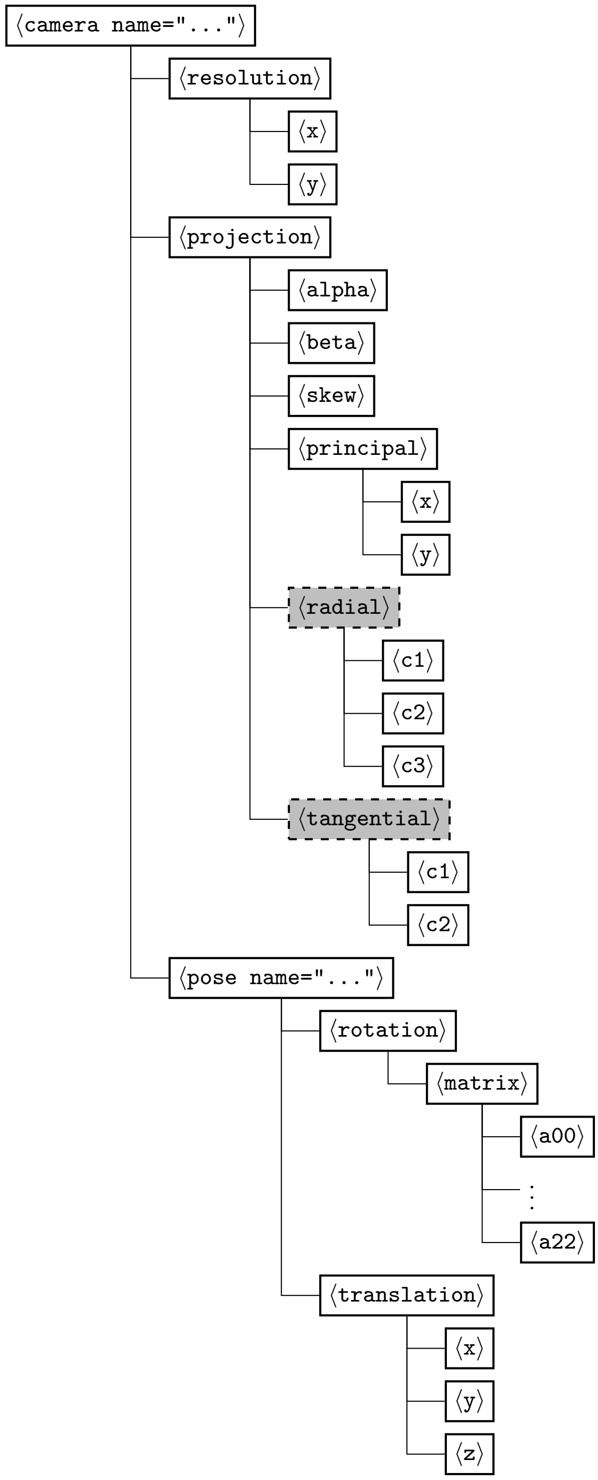

Figure 1: XML diagram of a minimal camera model specification. Dashed items are optional.

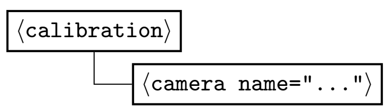

Figure 2: XML diagram of a multiple view setup specification.