Bundle Adjustment Based Multi-Camera Rig Calibration

Sileye Ba

June 29th 2016

1 Introduction

This document presents the bundle adjustment solution to the calibration of a synchronized multi-camera rig.

We consider a multi-camera rig composed of K cameras. Every camera k has intrinsic parameters $\theta_k = (f_k, c_k, d_k)$ (focal, optical center, distortion).

We denote the rigid motion to map a point in the world coordinate system to the k'th camera coordinate system as ${}^k\mathcal{T}_w = [{}^kR_w \; {}^kT_w]$, where ${}^kR_w$ is a rotation matrix, and ${}^kT_w$ a translation vector.

2 The Bundle Adjustment Problem

The camera rig has acquired t=1,...,T views. At each time t, camera k, acquires a view I_{kt}. At each time t, a number of N_{kt} keypoint, $x_{nkt},\; n = 1, ..., N_{kt}$ are detected on view I_{kt}.

Using keypoints descriptor[1], keypoints x_{nkt} on camera k, can be matched to keypoints x_{mlt} on camera l, with a matching weights w_{nmklt}[2]. This means that, with confidence w_{nmklt}, x_{nkt} and x_{mlt} represent the projection of the same 3D world's point.

We denote ${}^wX_{nmt}$, the 3D point in the world coordinate system that projects as x_{nkt} onto camera k's and as x_{mlt} onto camera l's view. This is formally written as

$$\begin{aligned} x_{nkt} &= \Pi(\theta_k, {}^k\mathcal{T}_w {}^wX_{nmt}) \\ x_{mlt} &= \Pi(\theta_l, {}^l\mathcal{T}_w {}^wX_{nmt}) \end{aligned} \tag{1}$$

where $\Pi(\theta_k, {}^k\mathcal{T}_w .)$ is the projective transformation mapping 3D points represented into the world referential, into 2D points on the k'th camera image plane.

Thus, using relations in equation (1), to find the camera intrinsic parameters θ, the 3D world to cameras transforms $\mathcal{T}_w$, and the 3D world's points ${}^wX$, the bundle adjustment problem can be stated as finding the arg-max of the cost function:

$$J(\theta, \mathcal{T}_w, {}^wX) = \sum_t \sum_{k \neq l} \sum_{n,m} w_{nmklt} \left(||x_{nkt} - \Pi(\theta_k, {}^k\mathcal{T}_w {}^wX_{nmt})||_2^2 + ||x_{mlt} - \Pi(\theta_l, {}^l\mathcal{T}_w {}^wX_{nmt})||_2^2 \right) \tag{2}$$

where $||.||_2$ is the $\mathbb{R}^2$ Euclidean distance.

Let us denote by $\hat{x}_{nkt} = \Pi(\theta_k, {}^k\mathcal{T}_w {}^wX_{nmt})$, the image point obtained by projecting the 3D point X_{nmt}. Using this notation, the bundle adjustment problem can be written as

[1]In our case we use AKAZE descriptors.

[2]In a hard assignment matching situation, which is our case, w_{nmklt} is either 0 or 1.

$$J(\theta, \mathcal{T}_w, {}^wX) = \sum_t \sum_{k \neq l} \sum_{n,m} w_{nmklt} \left(||x_{nkt} - \hat{x}_{nkt}||_2^2 + ||x_{mlt} - \hat{x}_{mlt}||_2^2 \right) \tag{3}$$

3 Solving The Bundle Adjustment Problem

The bundle adjustment,the BA problem requires finding the parameters values minimizing problem (3). This can be done by finding parameters vanishing the gradient of cost function $J(\theta, \mathcal{T}_w, {}^wX)$ which involves three main terms: $\nabla_{\theta_k} J$, $\nabla_{\omega_k} J$ where ω_k is the special Euclidean rigid motion parametrization ${}^k\mathcal{T}_w$, and $\nabla_{X_{nmt}} J$.

If we denote by $J_{nm}(\theta_k, \omega_k, X_{nmt}) = ||x_{nkt} - \hat{x}(\theta_k, \omega_k, X_{nmt})||_2^2$ we need to compute differential of J_{nm} suffices to obtain the differential of J.

4 Differentials of J_{nm}

Differential of J_{nm} can is defined as:

$$\lim_{\beta \to 0} \frac{J_{nm}(\theta_k, \omega_k, X_{nmt}) + J_{nm}(\theta_k + \beta d\theta_k, \omega_k, X_{nmt})}{\beta} =< \nabla_{\theta_k} J_{nm}, d\theta_k > \tag{4}$$

where $<, >$ is a scalar product. Assuming available a first order approximation of $\hat{x}(\theta_k + \beta d\theta_k, \omega_k, X_{nmt})$ around $\hat{x}(\theta_k, \omega_k, X_{nmt})$ as

$$\hat{x}(\theta_k + \beta d\theta_k, \omega_k, X_{nmt}) \approx \hat{x}(\theta_k, \omega_k, X_{nmt}) + \beta \frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial \theta_k} d\theta_k + o(\beta^2) \tag{5}$$

where $\frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial \theta_k}$ is the Jacobian of $\hat{x}(\theta_k, \omega_k, X_{nmt})$ w.r.t. θ_k. Using differential formula in (4) and the first order approximation in (4), the differential $\nabla_{\theta_k} J_{nm}$ is

$$\nabla_{\theta_k} J_{nm} = \frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial \theta_k}^\top (x_{nkt} - \hat{x}(\theta_k, \omega_k, X_{nmt})) \tag{6}$$

Following similar reasoning as previously, the other differential can be obtained as:

$$\nabla_{\omega_k} J_{nm} = \frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial \omega_k}^\top (x_{nkt} - \hat{x}(\theta_k, \omega_k, X_{nmt}))$$
$$\nabla_{X_{nmt}} J_{nm} = \frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial X_{nmt}}^\top (x_{nkt} - \hat{x}(\theta_k, \omega_k, X_{nmt}))$$

Computing differential requires the derivation of Jacobians $\frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial \theta_k}$, $\frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial \omega_k}$, $\frac{\partial \hat{x}(\theta_k, \omega_k, X_{nmt})}{\partial X_{nmt}}$.

A Projection of 3D world points onto the k'th camera image plane $\hat{x}_k = \Pi(\theta_k, {}^k\mathcal{T}_w {}^wX)$

The principle of the projection model is the following.

First the 3D world, using the world to camera mapping, point ${}^wX$ is transformed into the camera reference as

$${}^kX = {}^k\mathcal{T}_w {}^wX = \mathcal{T}(\omega_k) {}^wX.$$

Then, denoting the Euclidean distance in the suitable space as $||.||$, the camera point is mapped to the camera unit sphere as

$$ {}^{s}X = \frac{{}^{k}X}{||{}^{k}X||}. $$

Then the point of the camera unit sphere translated by unity in the z axis direction, and projected onto the camera image plane as

$$ {}^{k}\tilde{\mathbf{x}} = \begin{pmatrix} \frac{{}^{s}X^{1}}{{}^{s}X^{3}+1} \\ \frac{{}^{s}X^{2}}{{}^{s}X^{3}+1} \\ 1 \end{pmatrix} $$

If we define $r = ||{}^{k}\mathbf{x}||$, the radially distorted projection image point is given by

$$ {}^{d}\mathbf{x} = g({}^{k}\mathbf{x}, \mathbf{k}) = \left(\sum_{n=1}^{3} k_n r^n + 1 - \sum_{n=1}^{3} k_n \right) {}^{k}\mathbf{x}. $$

If $\mathbf{f} = (f_1, f_2)$ denotes the camera focal, and $\mathbf{c} = (c_1, c_2)$ its principal point, and the intrinsic camera matrix defined as

$$ \mathbf{K} = \begin{pmatrix} f_1 & 0 & c_1 \\ 0 & f_2 & c_2 \\ 0 & 0 & 1 \end{pmatrix}, $$

the image point in pixel is given as

$$ \tilde{y}_k = \mathbf{K}\,{}^{d}\tilde{\mathbf{x}}. $$

The full projection model can be written as

$$ \tilde{y}_k(\mathbf{f}, \mathbf{c}, \mathbf{k}, \omega_k, {}^{w}X) = \mathbf{K}(\mathbf{f}, \mathbf{c}) \left(\sum_{n=1}^{3} k_n r({}^{k}\mathbf{x}(\omega_k, {}^{w}X))^n + 1 - \sum_{n=1}^{3} k_n \right) {}^{k}\mathbf{x}(\omega_k, {}^{w}X) $$