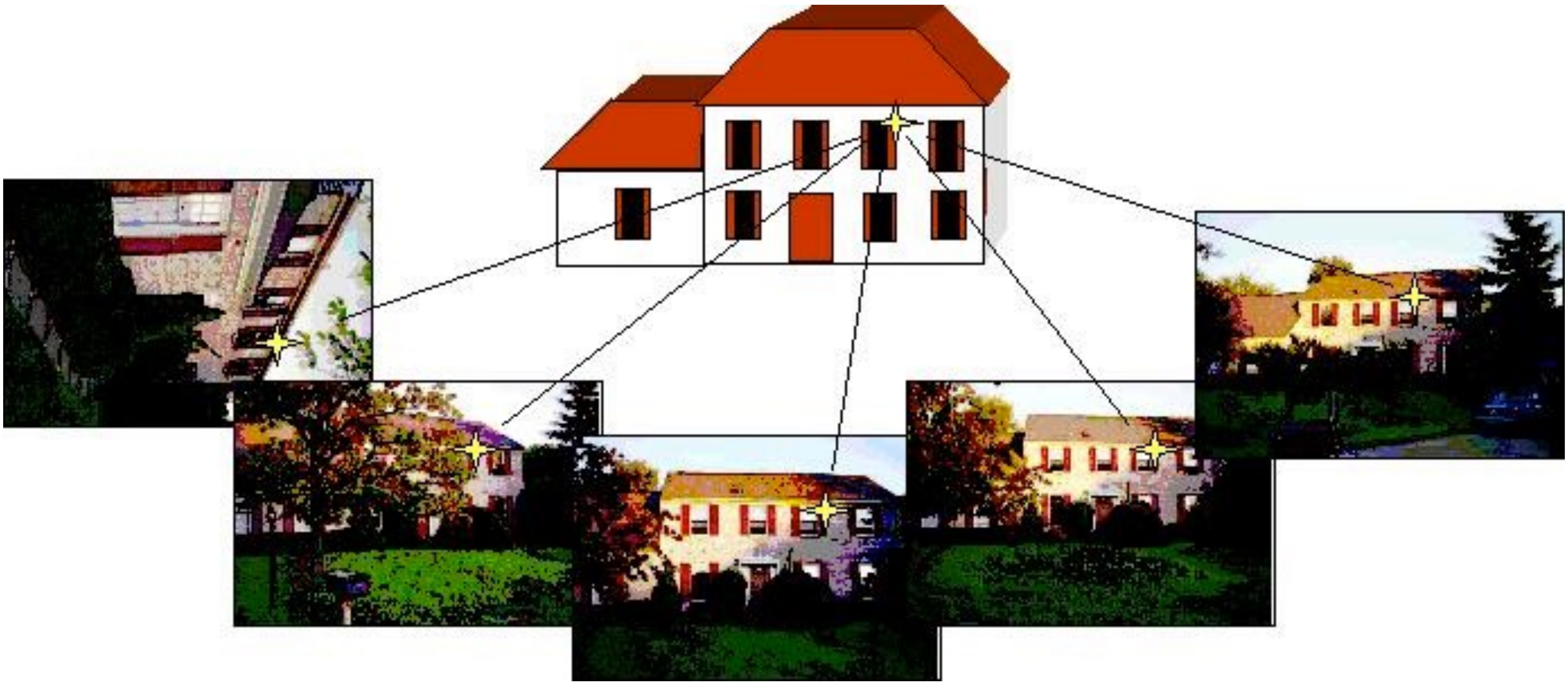
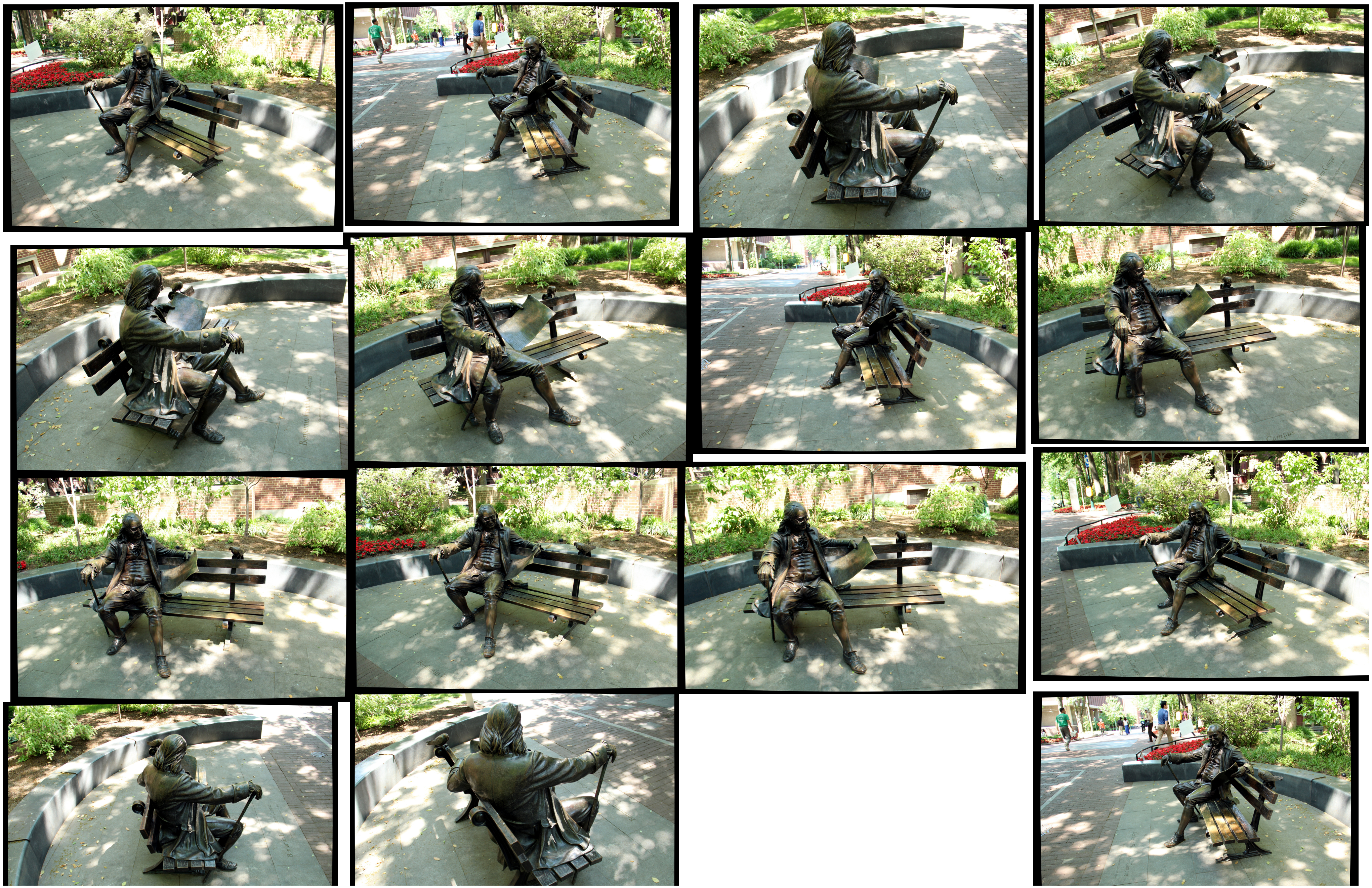
Perception: 3D Motion and Structure from Multiple Views or Bundle Adjustment

Kostas Daniilidis

Extract camera poses and structure from multiple views of the same scene



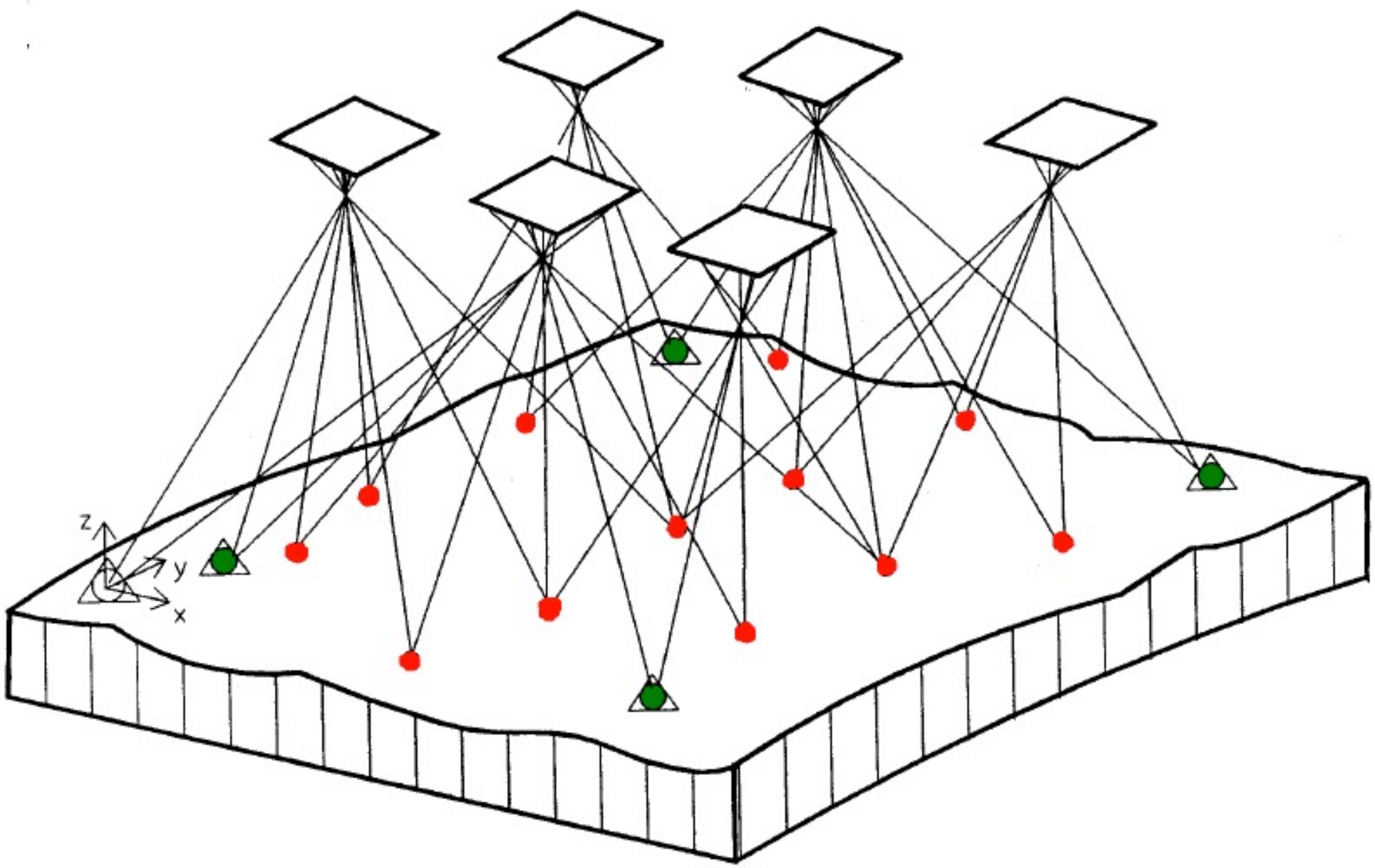
.. and an example closer to us



3D reconstruction

Urbanscape project 2006

„Bündelblockausgleichung“ is an old problem

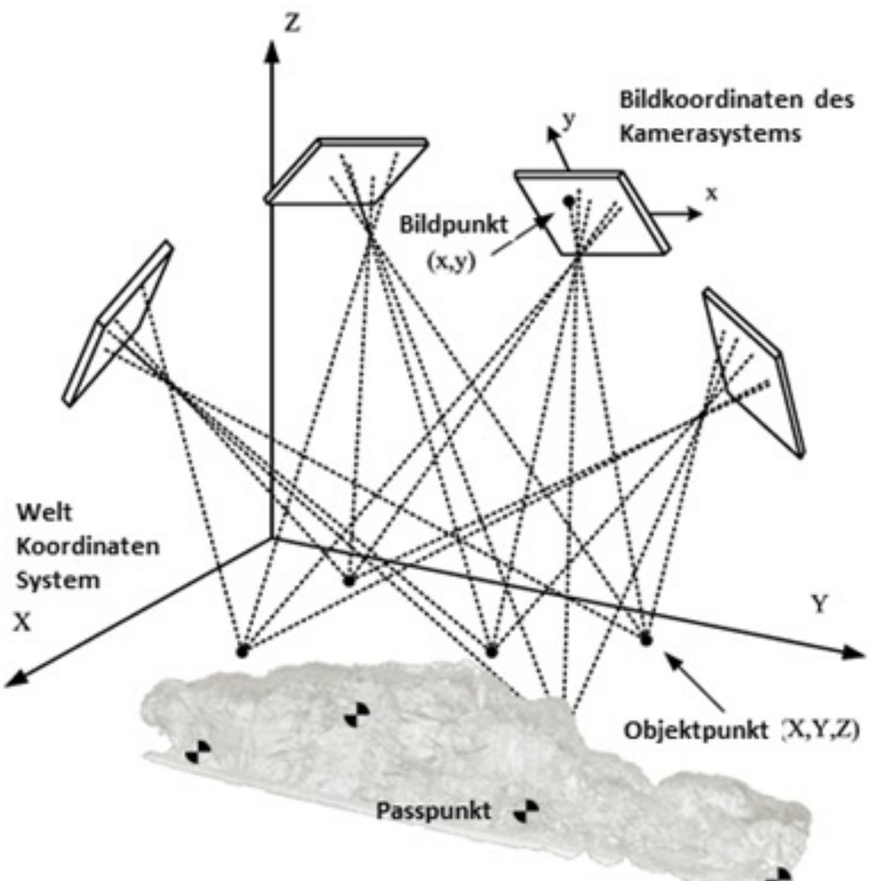


Some times as combination with PnP (resection) if ground control points (green) are known

Figure from photogeo.de

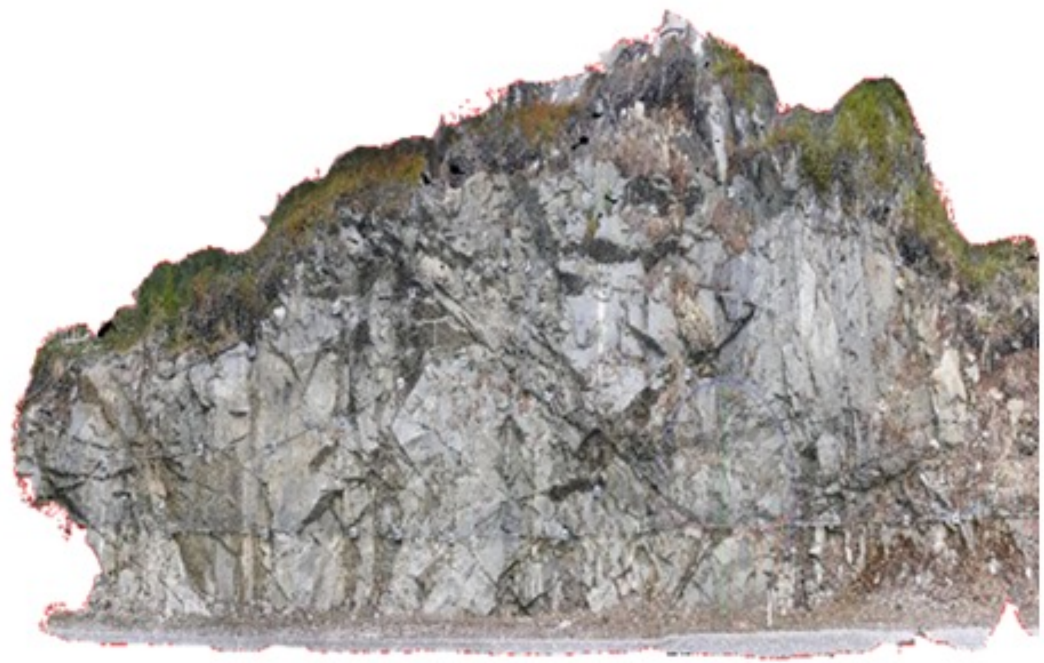
3D model from multiple views

3D-Geofotogrammetrische Aufnahme



Ergänzt nach: ajisaka.entopos.co.id

Ergebnis:
Entzerrtes und skalierbares 3D-Modell



Reference frame ambiguity hence we fix the first frame to be the world frame:

$$R_1 = I \qquad \text{and} \qquad T_1 = 0$$

Even with fixing the first frame, a global scale factor is still present. If we multiply all 3D points and T with the same scale measurements do not change.

Hence we have $\boxed{6(F-1)+3N-1}$ independent unknowns

and $2NF$ equations:

$$\begin{aligned} x_p^f &= \frac{R_{11}^f X_p + R_{12}^f Y_p + R_{13}^f Z_p + T_x}{R_{31}^f X_p + R_{32}^f Y_p + R_{33}^f Z_p + T_z} \\ y_p^f &= \frac{R_{21}^f X_p + R_{22}^f Y_p + R_{23}^f Z_p + T_y}{R_{31}^f X_p + R_{32}^f Y_p + R_{33}^f Z_p + T_z} \end{aligned}$$

If equations are independent (not always) then

$$\boxed{2NF \geq 6F + 3N - 7}$$

For two frames, it was already known that $N \geq 5$.

For three frames, $N \geq 4$.

Bundle Adjustment is the solution of this problem as nonlinear least-squares:

$$\arg\min_{R^f, T^f, X_p} \epsilon^T C^{-1} \epsilon$$

minimized with respect to all $6(F-1)$ motions and $3N-1$ structure unknowns, where ϵ is the error vector

$$\epsilon^T = \left(\cdots \quad \boxed{x_p^f - \frac{R_{11}^f X_p + R_{12}^f Y_p + R_{13}^f Z_p + T_x}{R_{31}^f X_p + R_{32}^f Y_p + R_{33}^f Z_p + T_z} \quad y_p^f - \frac{R_{21}^f X_p + R_{22}^f Y_p + R_{23}^f Z_p + T_y}{R_{31}^f X_p + R_{32}^f Y_p + R_{33}^f Z_p + T_z}} \quad \cdots \right.$$

and C is its error covariance. We will continue with the assumption that $C = I$.

Basics of nonlinear minimization

Call the objective function $\Phi(u) = \epsilon(u)^T \epsilon(u)$.

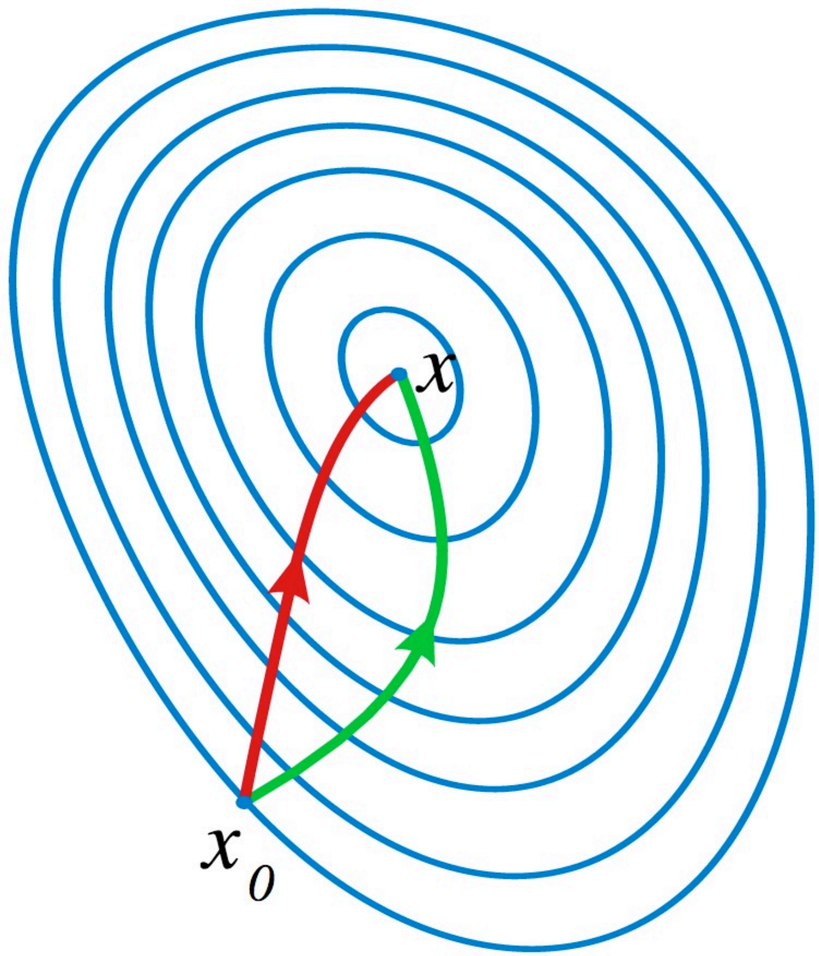
Given a starting value for the vector of unknowns u we iterate with steps Δu by locally fitting a quadratic function to $\Phi(u)$:

$$\Phi(u + \Delta u) = \Phi(u) + \Delta u^T \nabla \Phi(u) + \frac{1}{2} \Delta u^T H(u) \Delta u$$

where $\nabla \Phi$ is the gradient and H is the Hessian of Φ.

The minimum of this quadratic is at Δu satisfying

$$\boxed{H \delta u = -\nabla \Phi(u)}$$



The minimum of this quadratic is at Δu satisfying

$$\boxed{H\delta u = -\nabla\Phi(u)}$$

Vs the green gradient descent iteration.

If $\Phi(u) = \epsilon(u)^T \epsilon(u)$ then

$$\boxed{\nabla \Phi = 2 \sum_i \epsilon_i(u) \nabla \epsilon_i(u)^T = J(u)^T \epsilon}$$

where the Jacobian J consists of elements

$$J_{ij} = \frac{\partial \epsilon_i}{\partial u_j}$$

and the Hessian reads

$$H = 2 \sum_i \left(\nabla \epsilon_i(u) \nabla \epsilon_i(u)^T + \epsilon_i(u) \frac{\partial^2 \epsilon_i}{\partial u^2} \right) = 2 \left(J(u)^T J(u) + \sum_i \epsilon_i(u) \frac{\partial^2 \epsilon_i}{\partial u^2} \right)$$

by omitting quadratic terms inside the Hessian.

This yields the Gauss-Newton Iteration

$$\boxed{\Delta u = -(J^T J)^{-1} J^T \epsilon}$$

involving the inversion of a $(6F + 3N - 7) \times (6F + 3N - 7)$ matrix.

Bundle adjustment is about the “art” of inverting efficiently $(J^T J)$.

Let us split the unknown vector u into $u = (a, b)$ (following Sparse Bundle Adjustment paper by Spetsakis):

- $6F - 6$ motion unknowns a
- $3P - 1$ structure unknowns b.

We study the case of 4 points visible in 3 views.

We assume that the frames are 0,1,2 and that a_1 and a_2 are the 6x1 parameter vectors corresponding to transformations from 0 to 1 and from 0 to 2 respectively.

One point's depth is set to a fixed value to eliminate the scale ambiguity. We can do this by taking the calibrated coordinates and multiplying by a fixed depth.

The rest of the 3 points are b_1, b_2, b_3.

The Jacobian for 2 frames and 3 points has 6 pairs of rows (one pair for each image projection) and 15 columns/unknowns: columns/unknowns:

$$J = \frac{\partial \epsilon}{\partial(a,b)} = \begin{pmatrix} \underbrace{\begin{matrix} A_1^1 & 0 \\ 0 & A_1^2 \\ A_2^1 & 0 \\ 0 & A_2^2 \\ A_3^1 & 0 \\ 0 & A_3^2 \end{matrix}}_{\text{motion}} & \underbrace{\begin{matrix} B_1^1 & 0 & 0 \\ B_1^2 & 0 & 0 \\ 0 & B_2^1 & 0 \\ 0 & B_2^2 & 0 \\ 0 & 0 & B_3^1 \\ 0 & 0 & B_3^2 \end{matrix}}_{\text{structure}} \end{pmatrix}$$

with A matrices being 2×6 and B matrices being 2×3 being Jacobians of the error ϵ_i^f of the projection of the i-th point in the f-th frame.

We observe now a pattern emerging

$$J^T J = \begin{pmatrix} U^1 & 0 & W_1^1 & W_2^1 & W_3^1 \\ 0 & U^2 & W_1^2 & W_2^2 & W_3^3 \\ .. & .. & V_1 & 0 & 0 \\ .. & .. & 0 & V_2 & 0 \\ .. & .. & 0 & 0 & V_3 \end{pmatrix}$$

with the block diagonals for motion and structure separated.

We observe now a pattern emerging

$$J^T J = \begin{pmatrix} U^1 & 0 & W_1^1 & W_2^1 & W_3^1 \\ 0 & U^2 & W_1^2 & W_2^2 & W_3^3 \\ .. & .. & V_1 & 0 & 0 \\ .. & .. & 0 & V_2 & 0 \\ .. & .. & 0 & 0 & V_3 \end{pmatrix}$$

with the block diagonals for motion and structure separated.

Let us rewrite the basic iteration

$$(J^T J)\Delta u = -J^T \epsilon$$

as

$$\begin{pmatrix} U & W \\ W^T & V \end{pmatrix} \begin{pmatrix} \Delta a \\ \Delta b \end{pmatrix} = \begin{pmatrix} \epsilon'_a \\ \epsilon'_b \end{pmatrix}$$

and premultiply with

$$\begin{pmatrix} I & WV^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} U & W \\ W^T & V \end{pmatrix} \begin{pmatrix} \Delta a \\ \Delta b \end{pmatrix} = \begin{pmatrix} I & WV^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} \epsilon'_a \\ \epsilon'_b \end{pmatrix}$$

Motion parameters can be updated separately by inverting a $6F \times 6F$ matrix:

$$(U - WV^{-1}W^T)\Delta a = \epsilon'_a - WV^{-1}\epsilon'_b$$

Each 3D point can be be updated separately by inverting a $3\times$ 3 matrix V:

$$\boxed{V\Delta b = \epsilon'_b - W^T\Delta a}$$

If a point i does not appear in frame f then matrices A_i^f and B_i^f are set to zero.

Bundler©   Structure from Motion for Unordered Image Collections