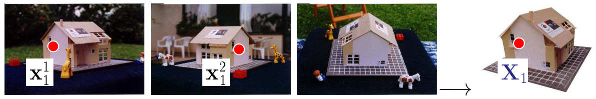
Structure from Many Perspective Images with Occlusions

Daniel Martinec Tomáš Pajdla

Center for Machine Perception, Czech Technical University in Prague

http://cmp.felk.cvut.cz

3D-reconstruction from perspective images with occlusions by factorization



Previous work: Tomasi & Kanade [1], Jacobs [3], Sturm & Triggs [2]

Main Contribution

Generalization of Jacobs [3] and combination with Sturm & Triggs [2] to obtain 3D-reconstruction from perspective images with occlusions by factorization

Problem Formulation

Perspective camera projection: $\lambda_p^i \underbrace{\mathbf{x}_p^i}_{3\times 1} = \underbrace{\mathrm{P}^i}_{3\times 4} \underbrace{\mathbf{X}_p}_{4\times 1}$

$$\underbrace{\begin{bmatrix} \lambda_1^1\mathbf{x}_1^1 & \lambda_2^1\mathbf{x}_2^1 & \dots & \lambda_n^1\mathbf{x}_n^1 \\ \times & \lambda_2^2\mathbf{x}_2^2 & & \times \\ \vdots & & \ddots & \vdots \\ \lambda_1^m\mathbf{x}_1^m & \times & \dots & \lambda_n^m\mathbf{x}_n^m \end{bmatrix}}_{\substack{\mathtt{R} \\ \text{rescaled measurement matrix}}} = \underbrace{\begin{bmatrix} \mathrm{P}^1 \\ \mathrm{P}^2 \\ \vdots \\ \mathrm{P}^m \end{bmatrix}}_{\substack{3m\times 4 \\ \text{motion}}} \underbrace{[\mathbf{X}_1\mathbf{X}_2\dots\mathbf{X}_n]}_{\substack{4\times n \\ \text{structure}}}$$

New Algorithm

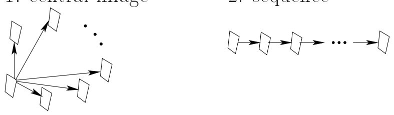
1. *Estimate (some)* λ_j^i using e.g. the epipolar geometry [2].
2. *Fill* $\times$ using the extension of [3] for perspective cameras.
   Repeat 1 and 2 until $\mathtt{R}$ is complete.
3. *Factorize* [1] complete $\mathtt{R}$.

Projective Depth Estimation (Sturm & Triggs [2])

uses fundamental matrices

$$\lambda_p^i = \frac{(\mathbf{e}^{ij} \wedge \mathbf{x}_p^i) \cdot (F^{ij}\ \mathbf{x}_p^j)}{\| \mathbf{e}^{ij} \wedge \mathbf{x}_p^i \|^2} \lambda_p^j \qquad \wedge \dots \text{cross-product}$$

- **Alternatives:** 1. central image 2. sequence



This work was supported by the grant CTU 0209313, the project GACR 102/00/1679, MSMT KONTAKT 2001/09 and ME412, and MSM 212300013.

New Filling for Perspective Cameras

- extension of filling for affine cameras ($\lambda_p^i = 1$) by Jacobs [3]
- fill $\mathtt{R}$ so that rank $\mathtt{R} = 4$
- points with unknown depth exploited

a 4-tuple of "LI" columns

$$\mathtt{B}_k = \begin{bmatrix} ?\ \mathbf{x}_1^1 & \lambda_2^1\mathbf{x}_2^1 & \lambda_3^1\mathbf{x}_3^1 & \lambda_4^1\mathbf{x}_4^1 \\ \lambda_1^2\mathbf{x}_1^2 & \lambda_2^2\mathbf{x}_2^2 & \lambda_3^2\mathbf{x}_3^2 & \lambda_4^2\mathbf{x}_4^2 \\ \vdots & & & \\ \lambda_1^m\mathbf{x}_1^m & \lambda_2^m\mathbf{x}_2^m & \lambda_3^m\mathbf{x}_3^m & \times \end{bmatrix} = \begin{bmatrix} ?\ x_1^1 & \lambda_2^1x_2^1 & \lambda_3^1x_3^1 & \lambda_4^1x_4^1 \\ ?\ y_1^1 & \lambda_2^1y_2^1 & \lambda_3^1y_3^1 & \lambda_4^1y_4^1 \\ ?\ w_1^1 & \lambda_2^1w_2^1 & \lambda_3^1w_3^1 & \lambda_4^1w_4^1 \\ \lambda_1^2x_1^2 & \lambda_2^2x_2^2 & \lambda_3^2x_3^2 & \lambda_4^2x_4^2 \\ \lambda_1^2y_1^2 & \lambda_2^2y_2^2 & \lambda_3^2y_3^2 & \lambda_4^2y_4^2 \\ \lambda_1^2w_1^2 & \lambda_2^2w_2^2 & \lambda_3^2w_3^2 & \lambda_4^2w_4^2 \\ \vdots & & & \\ \lambda_1^mx_1^m & \lambda_2^mx_2^m & \lambda_3^mx_3^m & \times \\ \lambda_1^my_1^m & \lambda_2^my_2^m & \lambda_3^my_3^m & \times \\ \lambda_1^mw_1^m & \lambda_2^mw_2^m & \lambda_3^mw_3^m & \times \end{bmatrix}$$

linear hull of all possible fillings

$$\mathcal{B}_k = Span\left(\begin{bmatrix} 0 & x_1^1 & \lambda_2^1x_2^1 & \lambda_3^1x_3^1 & \lambda_4^1x_4^1 & 0 & 0 & 0 \\ 0 & y_1^1 & \lambda_2^1y_2^1 & \lambda_3^1y_3^1 & \lambda_4^1y_4^1 & 0 & 0 & 0 \\ 0 & w_1^1 & \lambda_2^1w_2^1 & \lambda_3^1w_3^1 & \lambda_4^1w_4^1 & 0 & 0 & 0 \\ \lambda_1^2x_1^2 & 0 & \lambda_2^2x_2^2 & \lambda_3^2x_3^2 & \lambda_4^2x_4^2 & 0 & 0 & 0 \\ \lambda_1^2y_1^2 & 0 & \lambda_2^2y_2^2 & \lambda_3^2y_3^2 & \lambda_4^2y_4^2 & 0 & 0 & 0 \\ \lambda_1^2w_1^2 & 0 & \lambda_2^2w_2^2 & \lambda_3^2w_3^2 & \lambda_4^2w_4^2 & 0 & 0 & 0 \\ \vdots & & & & & & & \\ \lambda_1^mx_1^m & 0 & \lambda_2^mx_2^m & \lambda_3^mx_3^m & 0 & 1 & 0 & 0 \\ \lambda_1^my_1^m & 0 & \lambda_2^my_2^m & \lambda_3^my_3^m & 0 & 0 & 1 & 0 \\ \lambda_1^mw_1^m & 0 & \lambda_2^mw_2^m & \lambda_3^mw_3^m & 0 & 0 & 0 & 1 \end{bmatrix}\right)$$

- $\mathcal{B}_k$ contains filled $\mathtt{B}_k$ (also $\mathtt{R}$) $\longrightarrow$ constraint on $\mathtt{R}$
- intersection of many $\mathcal{B}_k$ $\longrightarrow$ filled $\mathtt{R}$

Heuristics guiding the filling of R

- many alternatives for computing λ_p^i
- noise $\longrightarrow$ alternatives not equivalent
- heuristics $\longrightarrow$ good reconstruction
- Fs not computed in heuristics

Principle 1 *The more depths known before the filling, the smaller the expected error.*

only Fs used: very much data missing $\longrightarrow$ depth estimation and filling must be repeated

Principle 2 *The more image points that are filled in one step, the smaller the expected error.*

References

[1] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. In *IJCV(9)2*, 1992.

[2] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *ECCV*, 1996.

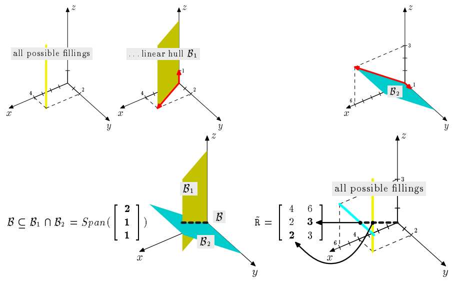
[3] D. Jacobs. Linear fitting with missing data: Applications to structure from motion and to characterizing intensity images. In *CVPR*, 1997.

Explanation of Filling on a Simplified Example

affine cameras, rank $\mathtt{R} = 1$

Fill $\mathtt{R} = \begin{bmatrix} 4 & 6 \\ 2 & \times \\ \times & 3 \end{bmatrix}$ so that rank $\mathtt{R} = 1$.

$\begin{bmatrix} 4 \\ 2 \\ \times \end{bmatrix} \dots \mathcal{B}_1 = Span\left(\begin{bmatrix} 4 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}\right)$, $\begin{bmatrix} 6 \\ \times \\ 3 \end{bmatrix} \dots \mathcal{B}_2 = Span\left(\begin{bmatrix} 6 & 0 \\ 0 & 1 \\ 3 & 0 \end{bmatrix}\right)$

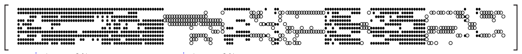


LM = lin. method, BA = bundle adjustment

Wide Base-Line Stereo

	Scene *House*	10 images [2952×2003]
	Point detection	manual, 203 points in space
	Depth estimation	central image No. 1
	Amount of missing data	47.83 %
LM	Mean error per image point [pxl]	3.91
LM + BA		1.44

$\mathtt{R} =$



"•" known λ_p^i (75.7 %), "∘" unknown λ_p^i (24.3 %), " " occlusion

Dense Sequence (Oxford)

	Scene *Dinosaur (Oxford)*	36 images [720×576]
	Point detection	Harris' operator, 4983 points in space
	Depth estimation	sequence
	Amount of missing data	90.84 %
LM	Mean error per image point [pxl]	1.76
LM + BA		0.64

$\mathtt{R} =$

"•" known λ_p^i (100.0 %), " " occlusion

Simulation: repr. error as a function of noise and missing data