Plane + Parallax, Tensors & Factorization

Bill Triggs

INRIA Rhône-Alpes

Grenoble, France

http://www.inrialpes.fr/movi/people/Triggs

Topics

1. Plane + parallax representation
2. Matching tensors in plane + parallax
3. Projective reconstruction by camera centre / projective height factorization.

Plane + Parallax

- The effective orientation and calibration of a camera can be changed arbitrarily by image warping.
- As warping changes nothing in 3D, the intrinsic properties (singularities, *etc.*) of visual reconstruction and matching geometry must be independent of orientation and calibration

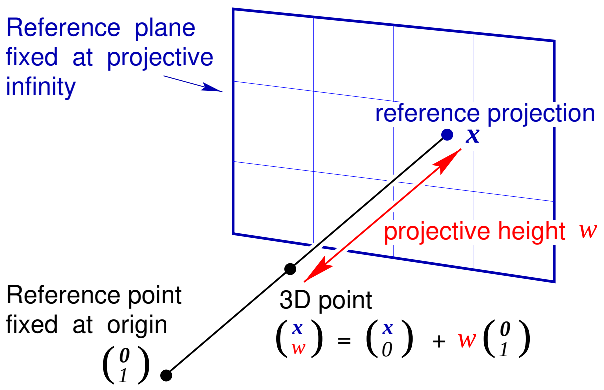
— *i.e.* they depend *only* on the camera centres.

- To make this explicit, “fixate” a given 3D ***reference plane***, warping all images so that the plane’s image is fixed.

— this greatly simplifies 3D and matching tensor calculations

— residual parallax displacements relative to the plane are often small, so feature detection and image processing are easier too.

‘Reference Plane at Infinity’ Coordinates



- This fixes 3D projective coordinates up to scale.
- Reference projection + height representation for 3D points: $\mathbf{X} = \binom{\boldsymbol{x}}{w}$

— w depends on normalization for $\boldsymbol{x}$.

Plane Aligned Cameras

- Plane-aligned cameras are parametrized by their projection centres:

$$\mathbf{P}_i \simeq (\boldsymbol{I}_{3\times 3} \,|\, -\boldsymbol{c}_i) \quad \text{with centre of projection} \quad \binom{\boldsymbol{c}_i}{1}$$

- $\boldsymbol{c}_i$ is the displacement of camera from the origin.

Plane aligned camera geometry is projectively equivalent to *translating calibrated cameras*

- This simplifies many formulae and helps intuition!

Point Projection and Parallax

- Each 3D point projects to its reference image, plus a ***parallax*** term:

$$\boxed{\lambda_{ip}\,\mathbf{x}_{ip} \;=\; \mathrm{P}_i \begin{pmatrix} \mathbf{x}_p \\ w_p \end{pmatrix} \;=\; \mathbf{x}_p - w_p\,\mathbf{c}_i}$$

- λ_{ip} is a '***projective depth***' of $\mathbf{x}_{ip}$ — an unknown homogeneous scale factor.
- ***Bilinearity***: $(\text{parallax})_{ip} \;=\; (\text{height } w_p)\,(\text{camera displacement } \mathbf{c}_i)$

— this suggests a ***factorization based reconstruction*** method

— but first we must recover the missing depths λ_{ip}.

- Similarly, for the parallax between two cameras:

$$\boxed{\lambda_{ip}\,\mathbf{x}_{ip} - \lambda_{jp}\,\mathbf{x}_{jp} \;=\; -w_p\,\mathbf{c}_{ij} \qquad\qquad \mathbf{c}_{ij} \;\equiv\; \mathbf{c}_i - \mathbf{c}_j}$$

Matching Tensors in Plane + Parallax

- For aligned projections, the matching tensors and their relations reduce to *very simple* functions of the relative camera positions $\boxed{\mathbf{c}_{ij} \;\equiv\; \mathbf{c}_i - \mathbf{c}_j}$:

$$\boxed{\begin{array}{rcll}
\mathbf{e}_{ij} & \simeq & \mathbf{c}_{ij} & \text{epipole of c.o.p. } i \text{ in image } j \\
\mathbf{F}_{ij} & \simeq & [\,\mathbf{c}_{ij}\,]_\times & \text{fundamental matrix} \\
\mathbf{T}_i^{\,jk} & \simeq & \mathbf{I}_i^j \otimes \mathbf{c}_{ik} - \mathbf{c}_{ij} \otimes \mathbf{I}_i^k & \text{trifocal tensor} \\
\mathbf{Q}^{A_1A_2A_3A_4} & \simeq & \sum_{i=1}^{3} (-1)^{i-1} \boldsymbol{\epsilon}^{A_1 \ldots \hat{A}_i \ldots A_4} \cdot \mathbf{c}_{i4}^{A_i} & \text{quadrifocal tensor}
\end{array}}$$

- These have simple geometric interpretations, *e.g.* $\mathbf{e}_{ij}$ is the intersection of the ray from $\mathbf{c}_j$ to $\mathbf{c}_i$ with the reference plane.
- The tensors have coherent, meaningful relative scalings given by the RHS.

Matching Tensors & Their Relations

Matching tensors $\mathbf{T}$

- Quartic functions of P's, obtained by eliminating their 3D dependencies.

Matching constraints

- Multilinear consistency relations between $\mathbf{T}$'s and unscaled image features.

Depth recovery relations

- Bilinear consistency relations between $\mathbf{T}$'s and *coherently scaled* image features.

Closure relations

- Bilinear consistency relations between $\mathbf{T}$'s and P's.

Grassmann relations

- Bilinear consistency relations on $\mathbf{T}$'s.

Matching Constraints in Plane + Parallax

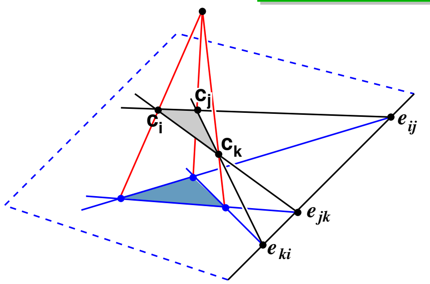
- Under plane alignment, the first few matching relations reduce to:

$$\boxed{\begin{array}{rcll}
[\,\mathbf{x}_i, \mathbf{e}_{ij}, \mathbf{x}_j\,] & = & 0 & \text{epipolar point} \\
(\mathbf{x}_i \wedge \mathbf{x}_j)\,(\mathbf{e}_{ik} \wedge \mathbf{x}_k)^\top - (\mathbf{e}_{ij} \wedge \mathbf{x}_j)\,(\mathbf{x}_i \wedge \mathbf{x}_k)^\top & = & \mathbf{0} & \text{trifocal point} \\
(\mathbf{l}_i \wedge \mathbf{l}_j)\,(\mathbf{l}_k \cdot \mathbf{e}_{ik}) - (\mathbf{l}_j \cdot \mathbf{e}_{ij})\,(\mathbf{l}_i \wedge \mathbf{l}_k) & = & \mathbf{0} & \text{trifocal line}
\end{array}}$$

- The trifocal point constraint contains two epipolar constraints, and also a proportionality-of-parallax relation.

Epipole Alignment and Desargues Configurations

- Cyclic sums of camera displacements vanish: $\boxed{\boldsymbol{c}_{ij} + \boldsymbol{c}_{jk} + \ldots + \boldsymbol{c}_{mi} \; = \; \boldsymbol{0}}$
- Given epipoles $\boldsymbol{e}_{ij}$, this can be used to reconstruct the $\boldsymbol{c}_{ij}$, *i.e.* to recover coherent scale factors.
- Circular triplets of epipoles are aligned: $\boxed{[\boldsymbol{e}_{ij}, \boldsymbol{e}_{jk}, \boldsymbol{e}_{ki}] \; = \; 0}$



- It is *much easier* to work with the elementary 3-vector geometry of the $\boldsymbol{c}$'s than the Desargues geometry of the $\boldsymbol{e}$'s.

Finding the Projective Depths

- To find the depths λ we use matching tensor based ***depth recovery relations***:

$$\begin{array}{rcll} \boldsymbol{F}_{ij}\,(\lambda_j\,\boldsymbol{x}_j) + \boldsymbol{e}_{ji} \wedge (\lambda_i\,\boldsymbol{x}_i) & = & \boldsymbol{0} & \text{epipolar} \\ \boldsymbol{T}_i^{\;jk}\,(\lambda_i \boldsymbol{x}_i) - (\lambda_j\,\boldsymbol{x}_j)\;(\boldsymbol{e}_{ik})^\top + \boldsymbol{e}_{ij}\;(\lambda_k\,\boldsymbol{x}_k)^\top & = & \boldsymbol{0} & \text{trifocal} \end{array}$$

- The matching tensors are estimated from the image correspondences as usual.

Projective SFM by Matrix Factorization

- Gather the homogeneous image projections $\lambda_{ip}\,\boldsymbol{x}_{ip} \; = \; \mathbf{P}_i\,\mathbf{X}_p$ of m general projective cameras $\mathbf{P}_i$ viewing n 3D points $\mathbf{X}_p$ into a big matrix:

$$\begin{pmatrix} \lambda_{11}\,\boldsymbol{x}_{11} & \ldots & \lambda_{1n}\,\boldsymbol{x}_{1n} \\ \vdots & \ddots & \vdots \\ \lambda_{m1}\,\boldsymbol{x}_{m1} & \ldots & \lambda_{mn}\,\boldsymbol{x}_{mn} \end{pmatrix} = \begin{pmatrix} \mathbf{P}_1 \\ \vdots \\ \mathbf{P}_m \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 & \ldots & \mathbf{X}_n \end{pmatrix}$$

- Any self-consistent set of projective depths λ contains an implicit projective reconstruction ("Joint image" viewpoint).
- SVD based ***rank 4 matrix factorization*** consolidates this and makes it explicit.

Plane Aligned Depth Recovery Relations

- Under reference plane alignment, the depth recovery relations become:

$$\begin{array}{rcll} \boldsymbol{e}_{ij} \wedge (\lambda_i\,\boldsymbol{x}_i - \lambda_j\,\boldsymbol{x}_j) & = & \boldsymbol{0} & \text{epipolar} \\ \boldsymbol{e}_{ij}\;(\lambda_k\,\boldsymbol{x}_k - \lambda_i\,\boldsymbol{x}_i)^\top - (\lambda_j\,\boldsymbol{x}_j - \lambda_i\,\boldsymbol{x}_i)\;(\boldsymbol{e}_{ik})^\top & = & \boldsymbol{0} & \text{trifocal} \end{array}$$

- These follow immediately from the relative parallax $\lambda_i\,\boldsymbol{x}_i - \lambda_j\,\boldsymbol{x}_j \; = \; -w\,\boldsymbol{c}_{ij}$ and $\boldsymbol{e}_{ij} \; \simeq \; \boldsymbol{c}_{ij}$.
- The trifocal relations contain two epipolar ones, plus an additional relative scale constraint.

Height / Camera Centre Factorization

- Plane alignment allows projective SFM factorization to be simplified.
- Rank 4 $(\text{projection}) \cdot (\text{structure})$ factorization becomes Rank 1 $(\text{camera centre}) \cdot (\text{height})$ factorization.

Inputs

- Feature correspondences under plane + parallax alignment.
- Estimates of the epipoles for projective depth recovery.

Output

- Projective 3D structure and camera centres, in a heuristic 'average of camera centres' frame.

Rationale

- For noiseless data, the parallax equations $\lambda_{ip}\,\boldsymbol{x}_{ip} \;=\; \boldsymbol{x}_p - w_p\,\boldsymbol{c}_i$ give:

$$\begin{aligned} \bar{\boldsymbol{x}}_p \;&\equiv\; \tfrac{1}{m}\textstyle\sum_{i=1}^m \boldsymbol{x}_{ip} \;\longrightarrow\; \boldsymbol{x}_p - w_p\,\bar{\boldsymbol{c}} & \qquad \bar{\boldsymbol{c}} \;&\equiv\; \tfrac{1}{m}\textstyle\sum_{i=1}^m \boldsymbol{c}_i \\ \boldsymbol{\delta x}_{ip} \;&\equiv\; \boldsymbol{x}_{ip} - \bar{\boldsymbol{x}}_p \;\longrightarrow\; w_p\,\boldsymbol{\delta c}_i & \qquad \boldsymbol{\delta c}_i \;&\equiv\; \boldsymbol{c}_i - \bar{\boldsymbol{c}} \end{aligned}$$

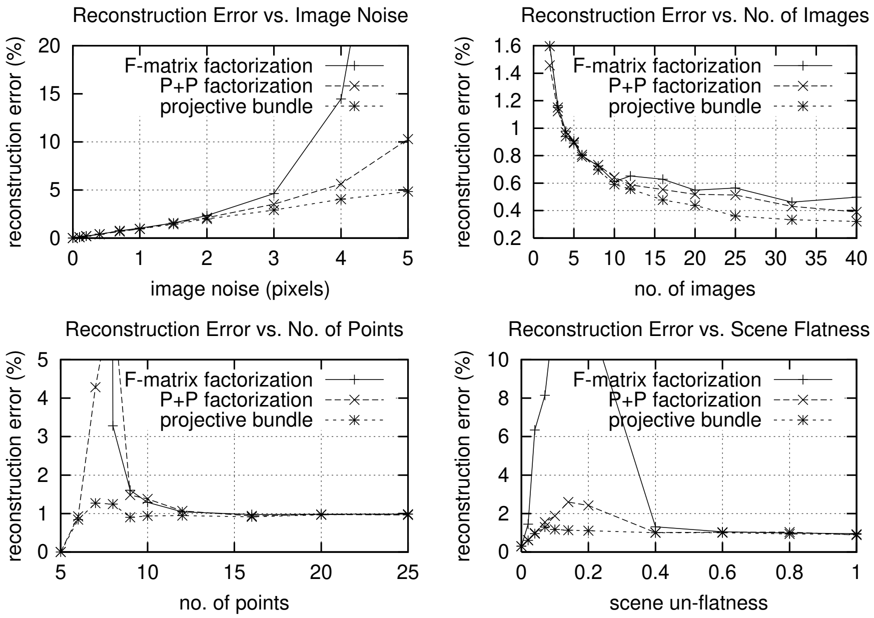
- $\bar{\boldsymbol{c}}$ is the same for all points and amounts to a change of origin.
- We can fix the origin freely, so we choose $\bar{\boldsymbol{c}} = \boldsymbol{0}$.

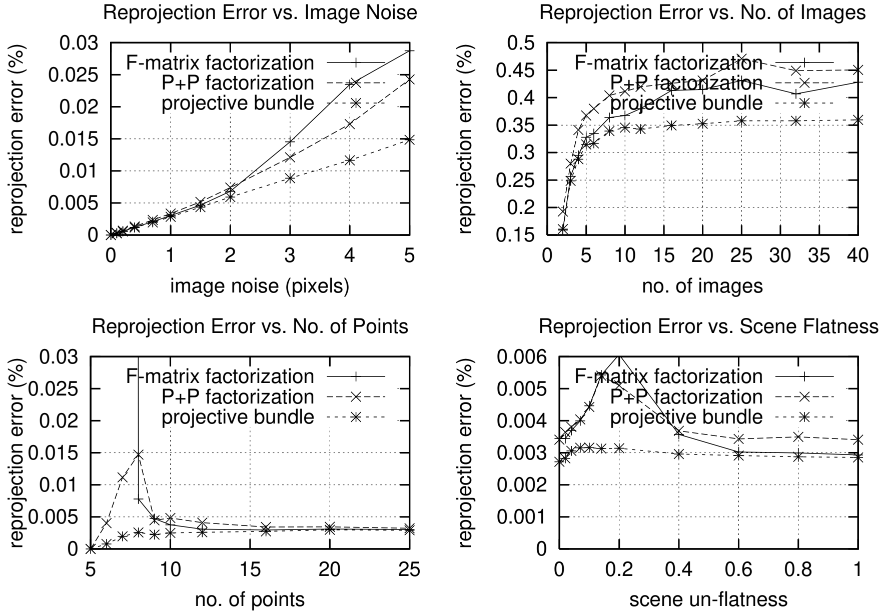
Algorithm

1. Find self-consistent projective depths λ_{ip} using epipolar or trifocal P+P depth recovery — as in the rank 4 method, a network of depth recovery relations are chained together.
2. Rescale the homogeneous image points: $\boldsymbol{x}_{ip} \;\leftarrow\; \lambda_{ip}\,\boldsymbol{x}_{ip}$
3. Centre the rescaled data: $\boldsymbol{\delta x}_{ip} \;\equiv\; \boldsymbol{x}_{ip} - \bar{\boldsymbol{x}}_p$ where $\bar{\boldsymbol{x}}_p \;\equiv\; \frac{1}{m}\sum_{i=1}^m \boldsymbol{x}_{ip}$
4. Factorize to rank 1 to give the projection centres and heights:

$$\begin{pmatrix} \boldsymbol{\delta x}_{11} & \dots & \boldsymbol{\delta x}_{1n} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\delta x}_{m1} & \dots & \boldsymbol{\delta x}_{mn} \end{pmatrix} \approx \begin{pmatrix} -\boldsymbol{c}_1 \\ \vdots \\ -\boldsymbol{c}_m \end{pmatrix} \begin{pmatrix} w_1 & \dots & w_n \end{pmatrix}$$

5. The final reconstruction is: $\mathbf{P}_i = (\boldsymbol{I}\,|\,-\boldsymbol{c}_i)$ and $\mathbf{X}_p \;=\; \binom{\bar{\boldsymbol{x}}_p}{w_p}$





Summary

1 Plane + parallax greatly simplifies matching tensor geometry and reconstruction.

2 Rather than working with epipole based Desargues constructions, it is easier to recover scale factors $\boldsymbol{e}_{ij} \longrightarrow \boldsymbol{c}_{ij}$ and use explicit 3-vector geometry.

3 Plane + parallax factorization looks promising, especially for near-planar scenes.

Extensions

• Other 3D features including lines and planes have simple plane + parallax representations.

• Both lines and planes can be added to the plane + parallax factorization

— for lines take via points

— for planes take homographies and treat their columns as point triplets.

• In both cases, the correct projective depths (homogeneous scalings) can be found by natural depth recovery methods.