Deep Fundamental Matrix Estimation
Supplemental Material

René Ranftl and Vladlen Koltun

Intel Labs

1 Network Architecture for Direct Regression

The architecture of the direct regression baseline network is shown in Table 1. It is based on the PoinNet architecture [2] to achieve permutation invariance. The architecture replaces the weighted least-squares layer with a max-pooling step over the feature maps along the point dimension, followed by a MLP.

Table 1. Network architecture for direct regression.

Layer	# in	# out	L-ReLU	Instance norm	Max pooling
1	–	64	✓	✓	✗
2	64	128	✓	✓	✗
3	128	1024	✓	✓	✗
4	1024	512	✓	✓	✗
5	512	256	✓	✓	✓
6	256	128	✓	✗	✗
7	128	64	✓	✗	✗
8	64	9	✓	✗	✗

2 Generating Virtual Correspondences

To generate the set of virtual correspondences we first define a regular grid over the image of size $M \times N$:

$$\mathbf{g}_{xy} = \Big(xN, yM\Big)^\top, \qquad x, y \in \{0, \delta, 2\delta, \dots, 1\}, \tag{1}$$

where $\delta = 0.01$ denotes the step size in the grid. Let $\mathbf{F}^{gt}$ denote the groundtruth fundamental matrix. We generate the set of virtual correspondences by projecting the points to the epipolar geometry using the Optimal Triangulation method [1]:

$$\begin{aligned} \tilde{\mathbf{p}}_i^{gt}, \tilde{\mathbf{p}}_i^{\prime gt} = &\underset{\mathbf{p}, \mathbf{p}^\prime \in \mathbb{R}^2}{\arg\min}\, d(\mathbf{g}_i, \mathbf{p})^2 + d(\mathbf{g}_i, \mathbf{p}^\prime)^2 \\ &\text{subject to } \hat{\mathbf{p}}^\top \mathbf{F}^{gt} \hat{\mathbf{p}}^\prime = 0, \end{aligned} \tag{2}$$

where $\hat{\mathbf{p}} = (\mathbf{p}^\top, 1)^\top$ denotes point $\mathbf{p}$ in homogeneous coordinates and $d(\mathbf{a}, \mathbf{b})$ denotes the geometric distance. We have that $\mathbf{p}_i^{gt} = (\tilde{\mathbf{p}}_i^{gt}, \tilde{\mathbf{p}}_i^{\prime gt}) \in \mathbb{R}^4$.

3 Homography estimation

The basis for homography estimation is formed by the Direct Linear Transform (DLT). Specifically, we have

$$(\mathbf{A}(\mathbf{P}))_{2i-1:2i} = \begin{pmatrix} -\hat{\mathbf{p}}_i^\top & \mathbf{0} & (\hat{\mathbf{p}}_i')_1 \hat{\mathbf{p}}_i^\top \\ \mathbf{0} & -\hat{\mathbf{p}}_i^\top & (\hat{\mathbf{p}}_i')_2 \hat{\mathbf{p}}_i^\top \end{pmatrix}, \quad g(\mathbf{x}) = (\mathbf{T}')^{-1}(\mathbf{x})_{3\times 3}\mathbf{T} = \mathbf{H}. \tag{3}$$

We use the symmetric transfer error for computing residuals and the loss:

$$r(\mathbf{p}_i, \mathbf{H}) = \left\| \frac{(\mathbf{H}\hat{\mathbf{p}}_i)_{1:2}}{(\mathbf{H}\hat{\mathbf{p}}_i)_3} - \mathbf{p}_i' \right\| + \left\| \frac{(\mathbf{H}^{-1}\hat{\mathbf{p}}_i')_{1:2}}{(\mathbf{H}^{-1}\hat{\mathbf{p}}_i')_3} - \mathbf{p}_i \right\|. \tag{4}$$

The training loss is again given as the mean clamped residual to the groundtruth correspondences of each stage, where groundtruth correspondences are generated by sampling a regular grid and distorting it according to the groundtruth homography.

4 Proof of Proposition 1

We need to solve the optimization problem

$$\begin{aligned} \mathbf{x}^{j+1} &= \underset{\mathbf{x}:\ \|\mathbf{x}\|=1}{\arg\min} \left\{ \left\| \mathbf{W}^j(\boldsymbol{\theta})\mathbf{A}\mathbf{x} \right\|^2 \right\} \\ &= \underset{\mathbf{x}:\ \|\mathbf{x}\|^2=1}{\arg\min}\ \left\| \mathbf{B}\mathbf{x} \right\|^2. \end{aligned} \tag{5}$$

To solve this problem, we form the Lagrangian:

$$\mathcal{L}(\mathbf{x}, \lambda) = \left\| \mathbf{B}\mathbf{x} \right\|^2 + \lambda(1 - \left\| \mathbf{x} \right\|^2) \tag{6}$$

The optimality conditions are

$$\mathbf{B}^\top \mathbf{B}\mathbf{x} - \lambda \mathbf{x} = 0 \tag{7}$$

$$1 - \left\| \mathbf{x} \right\|^2 = 0 \tag{8}$$

Rewriting (7) to

$$\mathbf{B}^\top \mathbf{B}\mathbf{x} = \lambda \mathbf{x} \tag{9}$$

implies that $\mathbf{x}$ is an Eigenvector of $\mathbf{B}^\top\mathbf{B}$, with associated Eigenvalue λ. It follows that

$$\mathbf{x}^{j+1} = \underset{\mathbf{x}:\ \|\mathbf{x}\|^2=1}{\arg\min}\ \left\| \mathbf{B}\mathbf{x} \right\|^2 \tag{10}$$

$$= \underset{\mathbf{x}:\ \|\mathbf{x}\|^2=1}{\arg\min}\ \mathbf{x}^\top \mathbf{B}^\top \mathbf{B}\mathbf{x} \tag{11}$$

$$= \underset{\mathbf{x}:\ \|\mathbf{x}\|^2=1}{\arg\min}\ \lambda \left\| \mathbf{x} \right\|^2 \tag{12}$$

Since $\|\mathbf{x}\| = 1$ by definition, we can see that (12) is minimized for the smallest eigenvalue λ_i. To see the connection to the singular value decomposition: Let $\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$. The columns of $\mathbf{V}$ correspond to the Eigenvectors of $\mathbf{B}^\top\mathbf{B}$, and their associated non-zero singular values are the square-roots of the non-zero Eigenvalues.

5 Failure Cases

Due to the structure of our approach, failure cases are similar to the baselines: Misidentificatication of inliers due to very high outlier ratios and inaccuracies due to nearly degenerate configurations of the inlier set. Examples of failure cases are shown in Figure 1.

Figure 1. Failure cases. Top row: First image with inliers (red) and outliers (blue). Bottom row: Epipolar lines of a random subset of groundtruth inliers in the second image. We show the epipolar lines of our estimate (green) and of the groundtruth (blue). Left: Misidentification of inliers. The bottom-most groundtruth inlier does not lie on its corresponding epipolar line. Middle: Failure in the very high noise regime. Right: Failure to pinpoint the exact epipolar geometry due to degenerate configuration. Images have been scaled and cropped for visualization.

6 Runtimes

Average runtimes for all evaluated approaches on the Tanks and Temples dataset are shown in Table 2. Note that MLESAC was evaluated using an unoptimized Python implementation.

Table 2. Comparison of runtimes.

	RANSAC	LMEDS	MLESAC	USAC	Ours
Time [ms]	11	12	696	24	26

References

1. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press (2000)
2. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep learning on point sets for 3D classification and segmentation. In: CVPR (2016)