Robust and Accurate Line- and Point-Based Pose Estimation without Manhattan Assumptions — Supplementary Material —

Yohann Salaün, Renaud Marlet, and Pascal Monasse

LIGM, UMR 8049, École des Ponts, UPEM, ESIEE Paris, CNRS, UPE, Champs-sur-Marne, France. Email: {yohann.salaun,renaud.marlet,pascal.monasse}@enpc.fr

We summarize here our settings for the different methods (Section 1) and we provide more details on our experiments (Section 2 and following).

1 Settings

1.1 Error measurements

The error e_R between an estimated rotation R and a ground truth rotation R_{gt} is computed with the following formula:

$$e_R = \angle(R^T R_{\text{gt}}) = \cos^{-1}\left(\frac{|\operatorname{Tr}(R^T R_{\text{gt}}) - 1|}{2}\right) \tag{1}$$

It combines both axis and angular errors for the rotation.

The error e_t between an estimated translation t and a ground truth translation t_{GT} is computed with the following formula:

$$e_t = \angle(t, t_{\text{gt}}) = \cos^{-1}\left(\frac{|t^T t_{\text{gt}}|}{\|t\|_2 \, \|t_{\text{gt}}\|_2}\right) \tag{2}$$

It measures the angular difference between the two directions, as the translation norm cannot be known.

1.2 RANSAC iterations

In our experiments, we assume a ratio of at least $w = 30\%$ of good matches for points and lines, which is sound after K-VLD filtering [1], and given that exact line matching is not required as long as a matched line segment is parallel to the actual corresponding line. At each RANSAC iteration, depending on the method, we pick at most $n = 6$ features. To be sure at confidence level $p = 95\%$ that we find at least one inlier model, at least $N_{\min} = \frac{\log(1-p)}{\log(1-w^n)} \approx 4108$ iterations are needed. The number of iterations to sample models in RANSAC is thus capped to $N = 5{,}000$.

More precisely, if the number of different sample combinations to considers (4-tuples or 5-tuples of points, 3-tuples or 2x2-tuples of lines, 2x2-tuples of lines plus 2-tuple of points) is lower than N, then all combinations are tried. Otherwise, only N are sampled. This makes sense for scenes with a low number of features, and in particular for lines in low-texture scenes. Note however that it only concerns model sampling; inliers are sought among all existing features.

1.3 Parameters

In all our experiments, parameters (when needed) are set as follows:

- Vanishing points, with their corresponding support lines, are merged after detection [2] using an incremental greedy strategy if their vanishing direction is the same up to a threshold of $5°$ (cf. Section 7 of the paper). This is enough to also prevent degenerate cases of pairs of parallel lines in methods 2x2-lines, mixed and AC-mixed (cf. Section 3.2 of the paper).
- For the 3-line method [3], we use the same thresholds as defined by the authors in their code. The RANSAC inlier threshold for rotation estimation, i.e., the maximum angle between corresponding vanishing directions, is set to $2°$, and the RANSAC inlier threshold for translation estimation, i.e., the maximum distance of a reprojected point to the epipolar line, is set to 0.2% of the image size.
- For methods 2x2-lines and mixed, the RANSAC inlier threshold for rotation estimation is set to the same value as in the original 3-line method [3], i.e., $2°$ (cf. Section 4.1 of the paper). It not only is a meaningful value; it also enables a fair comparison with the 3-line method.
  Moreover, the same angle is also used as RANSAC inlier threshold for the translation estimation, representing the maximum angle between epipolar planes. No pixel distance is required here (cf. Section 4.2 of the paper). This unification of initially heterogeneous parameters is made possible thanks to the new formulation of the translation error (cf. Eq. (5) of the paper).

These parameters are the same, for all the variety of scenes we consider.

The first parameter actually is just an add-on to the VP detector [2] to make sure that VP-based line clusters are separate enough. It has a status not much different than the parameters used by the line and point detectors and matchers, which are not relevant here. (They are however set to their "default" value.) As a matter of fact, all tested methods have as input exactly the same line and possibly point detections, and exactly the same matches.

The only major parameters that are specific to robust pose estimation, given detected features, are the two RANSAC thresholds for defining rotation and translation inliers. Our 2x2-line and mixed method only need one such parameter. The AC-mixed method does not need any; it is parameterless.

2 Sensitivity to noise w.r.t. lines and points, jointly

As mentioned in Section 7.1 of the paper, it is not clear how to model a joint realistic noise for both lines and points, as it might differ in nature and intensity: a Gaussian noise on line end points may not be consistent with a Gaussian noise on detected points, with the same σ. It is why we chose not to present a mixed experiment in the paper. Yet, under the strong assumption that such a common model makes sense, we provide here graphs, that show a smooth degradation with increasing noise (see Figure 7).

3 Robustness and accuracy of pose estimation on real datasets

Average results on scenes of the different datasets are shown on Table 1.

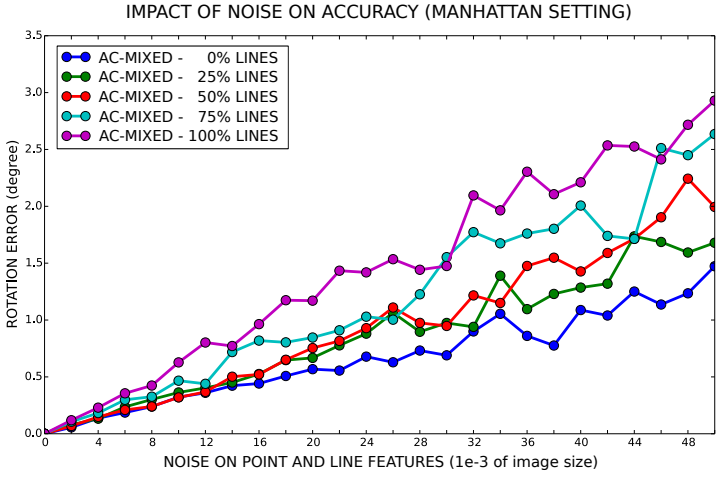

Fig. 7. Rotation error degrades smoothly as noise increases, for various points-lines ratios (same sampling setting as in Section 7.1).

Table 1. Average rotation and translation errors (in degrees) for the scenes of the real datasets. **Bold**: lowest errors (by 0.01°). Red: errors higher than 5°.

Method \ Dataset		Castle P18	Castle P30	Entry P10	Fountain P11	Herz-Jesu P8	Herz-Jesu P25	Office P8	Building P6	Car P4
Best VP	e_R	1.84	1.23	0.57	3.17	0.49	0.43	0.65	0.54	14.93
5-points [4]	e_R	**0.05**	**0.02**	**0.01**	**0.01**	**0.00**	**0.01**	6.88	0.23	**0.19**
	e_t	**0.31**	**0.29**	**0.12**	**0.09**	**0.11**	**0.13**	27.19	0.31	**0.20**
4-points	e_R	2.13	1.99	**0.02**	0.07	**0.00**	0.08	11.88	**0.19**	2.23
	e_t	13.23	14.26	0.14	0.29	0.15	0.28	18.68	**0.26**	16.03
3-lines [3]	e_R	0.30	1.07	0.23	0.84	0.24	0.31	6.45	6.68	24.25
	e_t	1.69	5.15	1.02	8.47	1.06	3.73	20.38	37.63	69.47
3-lines + SIFT	e_R	0.08	0.16	**0.01**	**0.01**	**0.00**	0.05	3.67	3.73	13.81
	e_t	0.42	0.87	0.17	0.18	0.18	0.34	16.26	18.72	30.46
2x2-lines	e_R	0.28	0.38	0.14	0.34	0.12	0.21	1.03	0.49	2.37
	e_t	1.32	1.71	0.63	0.98	0.63	0.92	3.26	1.57	18.03
mixed	e_R	0.22	0.27	0.14	0.30	0.06	0.16	1.01	0.24	0.75
	e_t	1.01	1.23	0.67	0.86	0.38	0.65	3.13	0.83	0.89
AC-mixed	e_R	**0.06**	0.05	**0.01**	**0.01**	**0.00**	**0.01**	**0.77**	**0.20**	0.22
	e_t	0.34	0.41	**0.13**	**0.08**	0.14	**0.13**	**2.48**	0.53	0.30

4 Office

- Number of images: 8
- Average number of line matches: 38
- Average number of point matches: 27

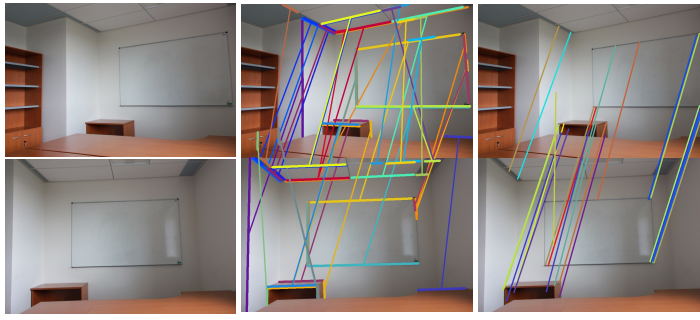

Fig. 8. Example of images from *Office*, line matches and points matches.

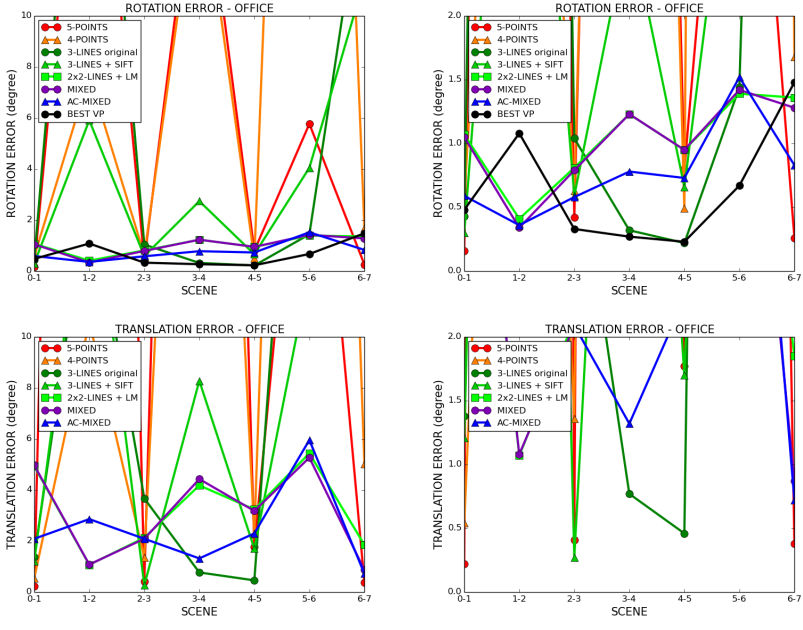

Fig. 9. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

5 Building

- Number of images: 6
- Average number of line matches: 75
- Average number of point matches: 192

Fig. 10. Example of images from *Building*, line matches and points matches.

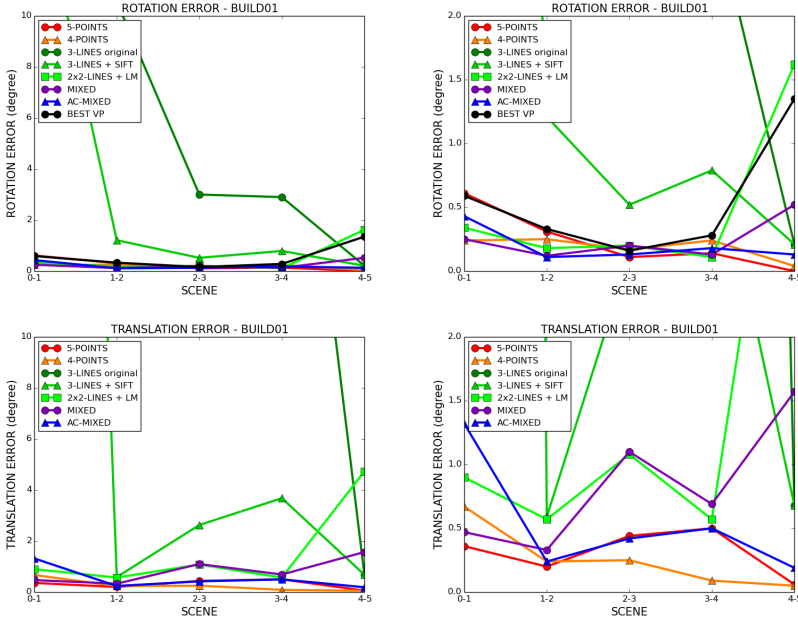

Fig. 11. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

6 Car

- Number of images: 4
- Average number of line matches: 39
- Average number of point matches: 1082

Fig. 12. Example of images from *Car*, line matches and points matches.

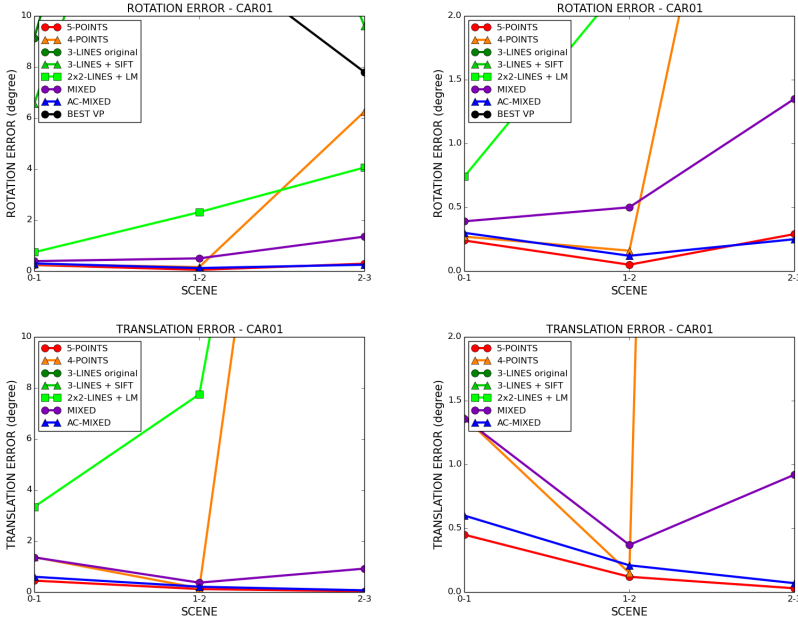

Fig. 13. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

7 Strecha - Castle P18

- Number of images: 18
- Average number of line matches: 158
- Average number of point matches: 2354

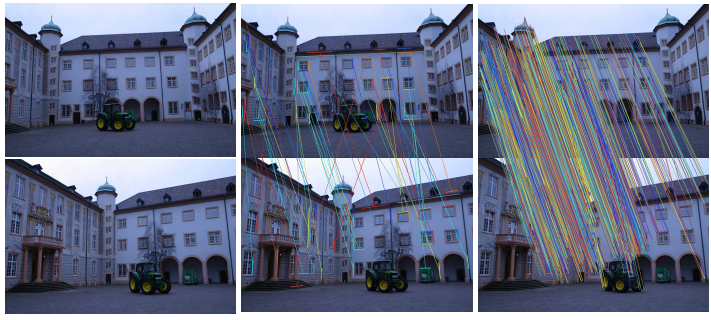

Fig. 14. Example of images from *Strecha - Castle P18 [5]*, line matches and points matches.

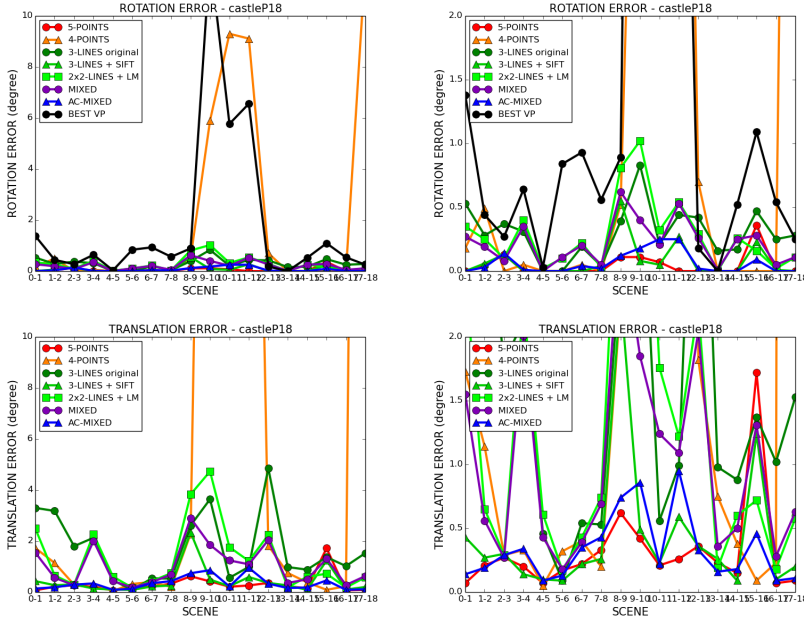

Fig. 15. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

8 Strecha - Castle P30

- Number of images: 30
- Average number of line matches: 173
- Average number of point matches: 2598

Fig. 16. Example of images from *Strecha - Castle P30 [5]*, line matches and points matches.

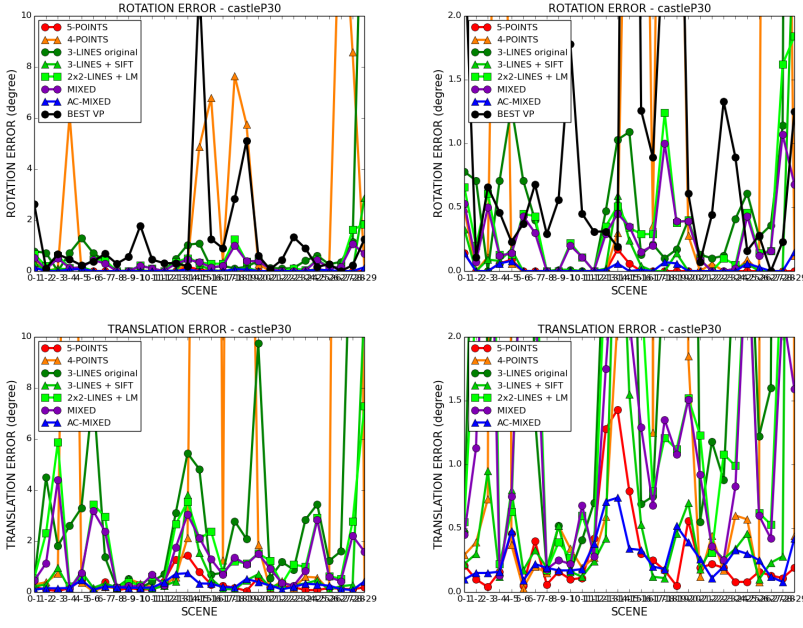

Fig. 17. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

9 Strecha - Entry P10

- Number of images: 10
- Average number of line matches: 143
- Average number of point matches: 2395

Fig. 18. Example of images from *Strecha - Entry P10 [5]*, line matches and points matches.

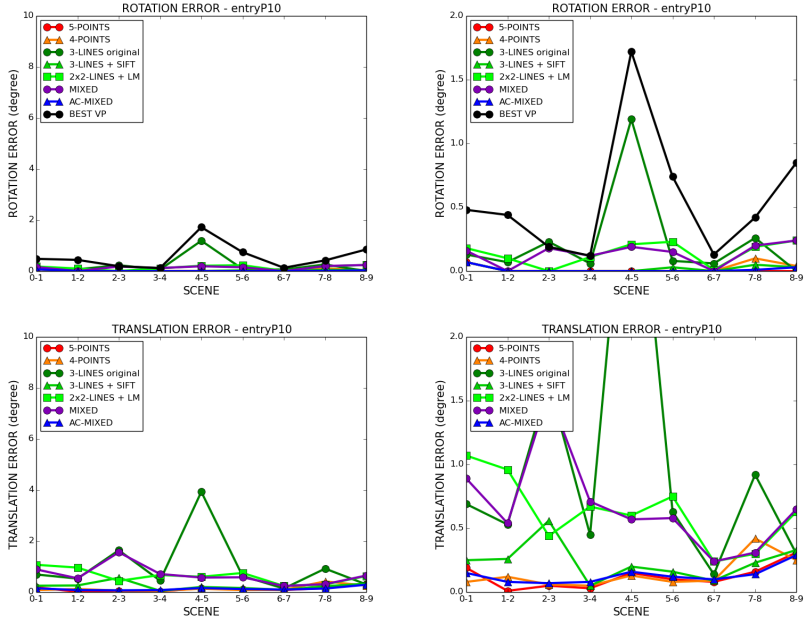

Fig. 19. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

10 Strecha - Fountain P11

- Number of images: 11
- Average number of line matches: 97
- Average number of point matches: 1854

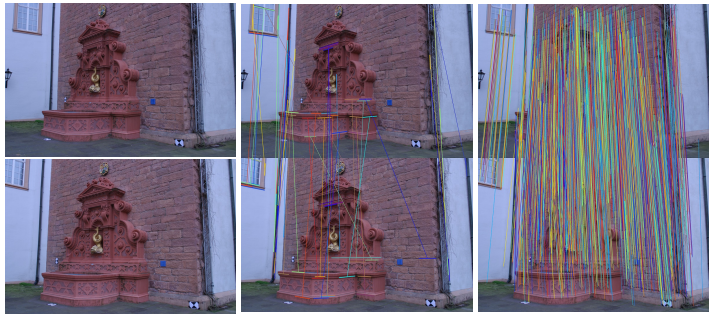

Fig. 20. Example of images from *Strecha - Fountain P11 [5]*, line matches and points matches.

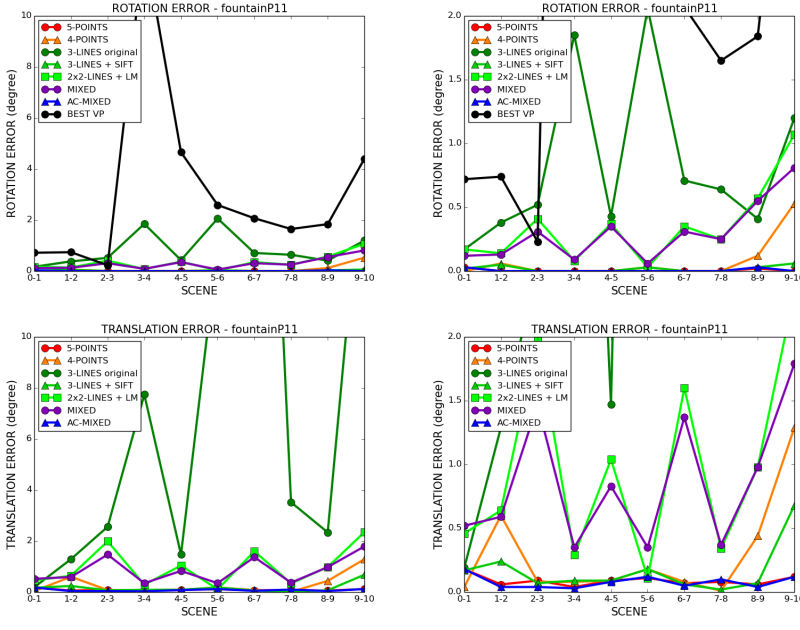

Fig. 21. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

11 Strecha - Herz Jesu P8

- Number of images: 8
- Average number of line matches: 129
- Average number of point matches: 1978

Fig. 22. Example of images from *Strecha - Herz Jesu P8 [5]*, line matches and points matches.

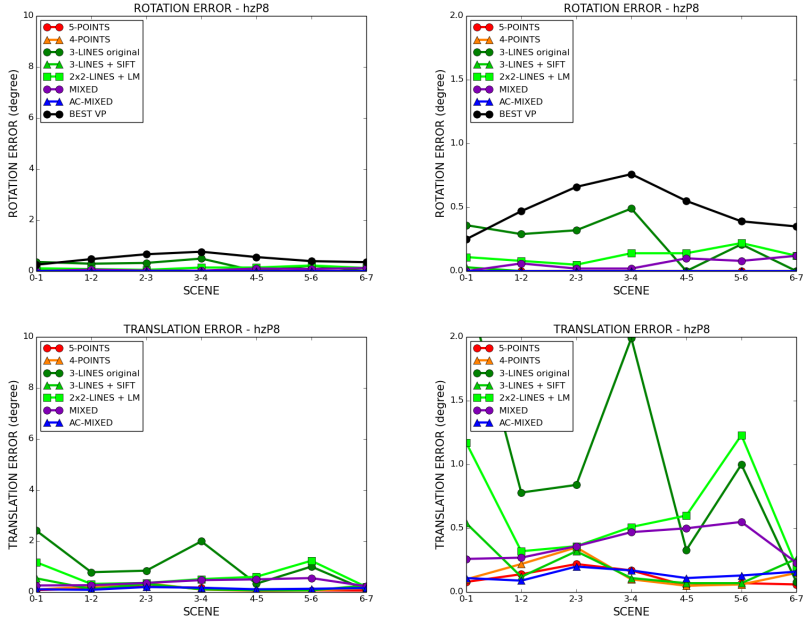

Fig. 23. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

12 Strecha - Herz Jesu P25

- Number of images: 25
- Average number of line matches: 124
- Average number of point matches: 1964

Fig. 24. Example of images from *Strecha - Herz Jesu P25 [5]*, line matches and points matches.

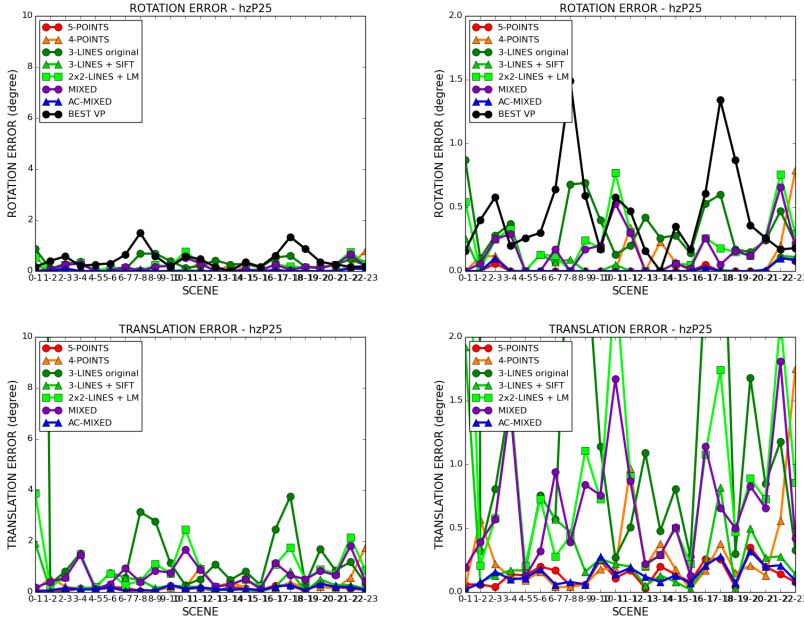

Fig. 25. Pose estimation error for each image pair. Top: rotation error. Bottom: translation error. Left: graph in range 0-10°. Right: zoomed graph in range 0-2°.

References

1. Liu, Z., Marlet, R.: Virtual line descriptor and semi-local matching method for reliable feature correspondence. In: British Machine Vision Conference (BMVC 2012), United Kingdom (September 2012) 16.1–16.11 1
2. Almansa, A., Desolneux, A., Vamech, S.: Vanishing point detection without any a priori information. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI 2003) **25**(4) (2003) 502–507 2
3. Elqursh, A., Elgammal, A.: Line-based relative pose estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011). (June 2011) 3049–3056 2, 3
4. Moulon, P., Monasse, P., Marlet, R.: Adaptive structure from motion with a contrario model estimation. In: Proceedings of the 11th Asian Conference on Computer Vision (ACCV 2012), Berlin, Heidelberg, Springer-Verlag (2013) 257–270 3
5. Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008). (June 2008) 1–8 7, 8, 9, 10, 11, 12