# Optimal Multi-view Correction of Local Affine Frames

Ivan Eichhardt
ivan.eichhardt@sztaki.mta.hu
Daniel Barath
barath.daniel@sztaki.mta.hu

Machine Perception Research Laboratory, MTA SZTAKI, Budapest, Hungary
Centre for Machine Perception, Department of Cybernetics Czech Technical University, Prague, Czech Republic

*[Supplementary Material]*

## 1 Additional synthetic experiments

To test the proposed method in a fully controlled environment, $N$ cameras were generated by their projection matrices looking towards the origin, each located in a random surface point on a sphere of radius 5. Then, a random 3D oriented point, at most one unit away from the origin and with random normal, was projected into the cameras. The ground truth LAF in each image was calculated from the projection matrix and the surface normal as in [3]. Zero-mean Gaussian noise with $\sigma$ standard deviation was added to both the point locations and affine parameters. Each reported result is averaged over $1,000$ runs.

In Fig. 1, the errors are plotted as the function of the noise level $\sigma$ (horizontal axis; in pixels). The shown values are: error of the noisy LAFs, *i.e.* the input without the correction (green curve), error of the corrected LAFs using the ground truth (red) and estimated (blue) fundamental matrices as input. The fundamental matrix was estimated from the noisy point correspondences using the normalised eight-point algorithm [5]. It can be seen that the more views are given, the better the corrected affine frames are. The proposed technique significantly improves the accuracy of the extracted LAFs even when the fundamental matrix is noisy as well.

Fig. 2 reports the mean, median and maximum processing times (in milliseconds) of the proposed method as the function of the view number. The values are calculated from all of the real-world experiments using our C++ implementation. It can be seen that the method is efficient, its processing time is negligible, *i.e.* $< 1$ millisecond, in most of the cases.

## 2 Additional evaluation of feature extractors

We present here the additional evaluation of several feature detectors, over the one presented in the paper. Commonly used feature extractors, listed below, are applied to images of the Strecha dataset [10] and their outputs are corrected by the proposed method. The dataset[1]



[1] Available at http://cvlabwww.epfl.ch/data/multiview/denseMVS.html

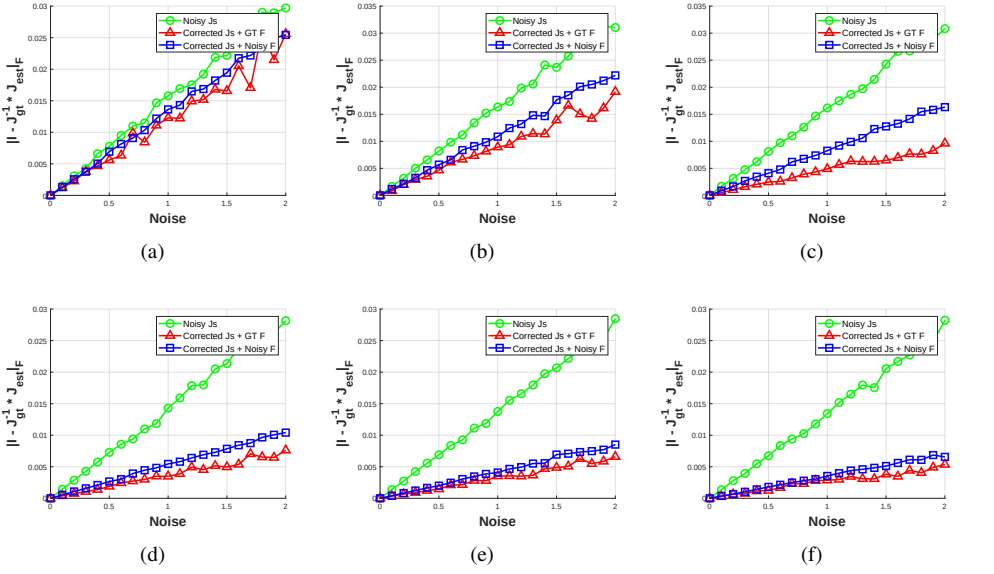


Figure 1: *Accuracy of the proposed method.* Diagrams (a-f) visualise results for 2, 3, 5, 10, 15 and 20 views, respectively. The error of the noisy and corrected LAFs are plotted as the function of the noise level $\sigma$ (horizontal axis; in pixels). The green curve shows the error of the input. For the red one, the ground truth fundamental matrix was used. For the blue curve, $\mathbf{F}$ was estimated from the noisy points. The error is $(1/K)\sum_{i=1}^{K}\left\|\mathbf{I}-\mathbf{M}_{i,\text{gt}}^{-1}\mathbf{M}_{i,\text{est}}\right\|_{\text{F}}$, where $K$ is the number of views, $\mathbf{I}$ is a $3\times 3$ identity matrix, $\mathbf{M}_{i,\text{gt}}$ and $\mathbf{M}_{i,\text{est}}$ are, respectively, the ground truth and estimated LAFs in the $i$th view.

consists of six image sequences of size $3072 \times 2048$ of buildings. Both the intrinsic and extrinsic parameters are given for all images.

To obtain ground truth LAFs in each image sequence, we first applied an SfM pipeline [8] with the known camera parameters obtaining a number of points along the images. Then, the points were manually assigned to dominant planes. Since each plane defines a homography between every view pair, the ground truth affine correspondences between the view pairs were calculated from the homography parameters as described in [2]. The evaluated extractors can be divided into three groups: (i) scale and rotation-covariant ones, like SIFT [6], AKAZE [1], Hessian [11], Difference of Gaussians (DoG) [11], and Harris-Laplace (Harris) [11]. (ii) Affine-covariant extractors using the Baumberg-iteration [4] such as Hessian-Aff, DoG-Aff and Harris-Aff, (iii) methods using simulated views, such as ASIFT [7], AAKAZE, ADoG, AHessian, and (iv) a shape-space based affine covariant region extractor TBMR [12] was also evaluated. In the experiments, the VlFeat library [11] provides the Hessian, DoG and Harries extractors, and their covariant counterparts: Hessian-Aff, DoG-Aff and Harris-Aff using its built-in version of the shape adaptation procedure (*i.e.*, the Baumberg iteration). We used the SIFT and AKAZE implementations included in OpenMVG [9]. For AAKAZE and ASIFT, the view-simulation of [7] is used, feeding warped versions of the input images to the detectors.

For the experiments, we used a modified version of OpenMVG [9] which, together with the point coordinates, stores the corresponding LAFs. For each detector, we performed fea-

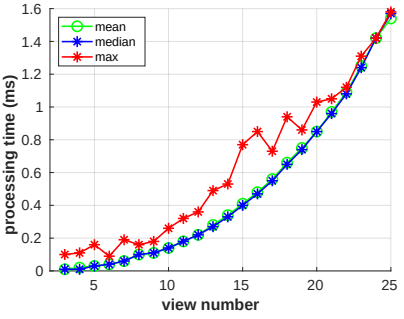


Figure 2: *The processing time* in milliseconds (mean, median and max) of the proposed method is plotted as the function of the view number. The values are calculated from the real-world experiments using our C++ implementation.

ture extraction, then established multi-view correspondences. The Global SfM pipeline [8] of OpenMVG estimated the camera motion and created a 3D point cloud of the scene. A robust triangulation procedure then established multi-view tracks of LAFs, with geometrically consistent centroids. Given the estimated poses, the matrix **B** of the epipolar constraints on LAFs was constructed and obtained LAFs were corrected by the proposed method.

The results are in Table 1. After the header, the odd rows report the accuracy of the extracted LAFs. The even rows show the quality of the corrected ones. Pairs of rows show the results of a particular detector. The sequences of the Strecha dataset are from the 3rd to 8th columns. The last two columns show the mean and median errors on the entire dataset.

It can be seen that the proposed method almost *always improved* the input LAFs. The most accurate detector is AAKAZE with the proposed correction. Also, the proposed technique significantly improves partially affine-covariant detectors, *e.g.* SIFT, as well.

Table 1 also contains the results *when using TBMR* [12], a shape-space based region detector. At first sight, seeing the large errors on average, one would think that is an inaccurate affine feature extractor. However, this case needs more thorough investigation since the median errors are the lowest among all the evaluated extractors. Consequently, at least the 50% of the extracted LAFs is accurate. The reason of this phenomenon is that the detector estimates the affine shape based on the concept of stable binary regions. For a region with low eccentricity and/or for a small area, the orientation of the underlying LAF becomes highly uncertain, thus, causing large errors which highly affects the average reported in Table 1. However, in other cases TBMR provides good LAFs with stable orientations.

# References

[1] P. Alcantarilla, J. Nuevo, and A. Bartoli. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In *Proc. British Machine Vision Conf.*, pages 13.1–13.11, Bristol, 2013. British Machine Vision Association. ISBN 978-1-901725-49-0. 00423.

[2] D. Barath and L. Hajder. A theory of point-wise homography estimation. *Pattern Recognition Letters*, 94:7 – 14, 2017. ISSN 0167-8655.

[3] D. Barath, J. Molnar, and L. Hajder. Novel methods for estimating surface normals from affine

| detector | LAF type | (a) | (b) | (c) | (d) | (e) | (f) | mean | median |
|---|---|---|---|---|---|---|---|---|---|
| AKAZE | Extracted | 0.23 | 0.22 | 0.27 | **0.27** | 0.30 | 0.26 | 0.26 | 0.20 |
| | Corrected | **0.12** | **0.12** | **0.14** | 0.62 | **0.18** | **0.17** | **0.22** | **0.09** |
| SIFT | Extracted | 0.22 | 0.22 | 0.23 | 0.26 | 0.31 | 0.29 | 0.26 | 0.20 |
| | Corrected | **0.14** | **0.12** | **0.13** | **0.18** | **0.18** | **0.21** | **0.16** | **0.11** |
| DoG | Extracted | 0.18 | 0.18 | 0.21 | 0.18 | 0.26 | 0.24 | 0.21 | 0.15 |
| | Corrected | **0.09** | **0.11** | **0.14** | **0.13** | **0.14** | **0.15** | **0.13** | **0.08** |
| Harris | Extracted | 0.25 | 0.25 | 0.26 | 0.25 | 0.30 | 0.28 | 0.27 | 0.21 |
| | Corrected | **0.15** | **0.14** | **0.13** | **0.16** | **0.17** | **0.19** | **0.15** | **0.10** |
| Hessian | Extracted | 0.25 | 0.25 | 0.26 | 0.26 | 0.33 | 0.29 | 0.27 | 0.22 |
| | Corrected | **0.14** | **0.14** | **0.12** | **0.16** | **0.20** | **0.20** | **0.16** | **0.11** |
| DoG-Aff | Extracted | 0.25 | 0.25 | **0.29** | 0.27 | 0.43 | 0.39 | 0.31 | 0.19 |
| | Corrected | **0.08** | **0.08** | 0.40 | **0.16** | **0.27** | **0.23** | **0.20** | **0.07** |
| Harris-Aff | Extracted | 0.30 | 0.30 | 0.31 | 0.38 | 0.40 | 0.35 | 0.34 | 0.25 |
| | Corrected | **0.14** | **0.13** | **0.13** | **0.24** | **0.16** | **0.18** | **0.16** | **0.10** |
| Hessian-Aff | Extracted | 0.29 | 0.29 | 0.29 | 0.37 | 0.41 | 0.35 | 0.33 | 0.25 |
| | Corrected | **0.13** | **0.12** | **0.13** | **0.23** | **0.16** | **0.18** | **0.16** | **0.10** |
| AAKAZE | Extracted | 0.26 | 0.25 | 0.31 | 0.32 | 0.30 | 0.28 | 0.29 | 0.22 |
| | Corrected | **0.11** | **0.10** | **0.13** | **0.18** | **0.12** | **0.13** | **0.13** | **0.08** |
| ADoG | Extracted | 0.20 | 0.21 | 0.25 | 0.28 | 0.26 | 0.26 | 0.24 | 0.17 |
| | Corrected | **0.09** | **0.11** | **0.13** | **0.16** | **0.12** | **0.13** | **0.12** | **0.07** |
| AHessian | Extracted | 0.27 | 0.25 | 0.29 | 0.28 | | 0.28 | 0.27 | 0.20 |
| | Corrected | **0.12** | **0.15** | **0.12** | **0.17** | | **0.16** | **0.14** | **0.08** |
| ASIFT | Extracted | 0.24 | 0.24 | 0.25 | 0.28 | 0.31 | 0.30 | 0.27 | 0.19 |
| | Corrected | **0.11** | **0.11** | **0.12** | **0.17** | **0.14** | **0.16** | **0.14** | **0.08** |
| TBMR (mean errors) | Extracted | **0.39** | 0.39 | 0.38 | **0.34** | **0.33** | **0.24** | **0.34** | 0.15 |
| | Corrected | 0.44 | **0.29** | **0.35** | 0.41 | 0.38 | 0.84 | 0.45 | **0.07** |
| TBMR (med. errors) | Extracted | 0.15 | 0.15 | 0.14 | 0.18 | 0.16 | 0.13 | | |
| | Corrected | **0.06** | **0.05** | **0.05** | **0.12** | **0.06** | **0.06** | | |

Table 1: *Comparison of feature detectors* in terms of the accuracy of the obtained LAFs. The accuracy (same metric as in Fig. 1) of the extracted and corrected (by the proposed method) LAFs are put in the odd and even rows, respectively. The scenes (columns) of the Strecha dataset: (a) `castle-P19`, (b) `castle-P30`, (c) `entry-P10`, (d) `fountain-P11`, (e) `herz-jesus-P25` and (f) `herz-jesus-P8` were fed into the [9] SfM pipeline. The proposed method almost *always improve* the extracted LAFs.

transformations. In *Proc. Joint Conf. on Computer Vision, Imaging and Computer Graphics Theory and Appl.* Springer International Publishing, 2016.

[4] A. Baumberg. Reliable feature matching across widely separated views. In *Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 774–781, Hilton Head Island, SC, USA, 2000. IEEE Comput. Soc. ISBN 978-0-7695-0662-3.

[5] R. I. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[6] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[7] J-M. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2009.

[8] P. Moulon, P. Monasse, and R. Marlet. Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Proc. International Conf. on Computer Vision*, pages 3248–3255, 2013.

[9] P. Moulon, P. Monasse, R. Perrot, and R. Marlet. OpenMVG: Open multiple view geometry. In *International Workshop on Reproducible Research in Pattern Recognition*, pages 60–74. Springer, 2016.

[10] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Conf. on Computer Vision and Pattern Recognition*. IEEE, 2008.

[11] A. Vedaldi and B. Fulkerson. VLFeat - an open and portable library of computer vision algorithms. In *Proc. ACM Conf. on Multimedia*, 2010.

[12] Y. Xu, P. Monasse, T. Géraud, and L. Najman. Tree-based morse regions: A topological approach to local feature detection. *IEEE Trans. Image Processing*, 23(12):5612–5625, 2014.