

# Approximate Models for Fast and Accurate Epipolar Geometry Estimation

James Pritts, Ondřej Chum and Jiří Matas

Center for Machine Perception

Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Cybernetics

Email: {prittjam,chum,matas}@cmp.felk.cvut.cz

**Abstract**—This paper investigates the plausibility of using approximate models for hypothesis generation in a RANSAC framework to accurately and reliably estimate the fundamental matrix. Two novel fundamental matrix estimators are introduced that sample two correspondences to generate affine-fundamental matrices for RANSAC hypotheses. A new RANSAC framework is presented that uses local optimization to estimate the fundamental matrix from the consensus correspondence sets of verified hypotheses, which are approximate models. The proposed estimators are shown to perform better than other approximate models that have previously been used in the literature for fundamental matrix estimation in a rigorous evaluation. In addition the proposed estimators are over 30 times faster, in terms of models verified, than the 7-point method, and offer comparable accuracy and repeatability on a large subset of the test set.

## I. INTRODUCTION

Epipolar geometry estimation is the initial step in many computer vision tasks, such as 3D-scene reconstruction, motion recovery, auto-calibration, and robot navigation. Feature correspondences are, in practice, corrupted by outliers (*i.e.* feature mismatches), which, if included in the model estimation, have an outsized and detrimental effect on model accuracy. While many approaches have been proposed to discard outliers, RANSAC and its many variants are typical and highly effective choices to obtain correct epipolar geometry for stereo problems [9, 11].

The expected run-time of RANSAC depends on two factors: the sample size required for generating model hypotheses and the proportion of correspondences that are outliers. Two methods have been proposed that estimate approximate models which require fewer correspondences than the 7-point method for RANSAC hypothesis generation. Goshen and Shimshoni [10] assume distinguished regions of SIFT correspondences are invariant up to a similarity and construct eight correspondences from two matching SIFT descriptors by using each SIFT's dominant gradient and scale [14]. This construction is used as input to the 8-point method in a RANSAC-like framework called Balanced Exploration and Exploitation Model Search (BEEM). Perdoch et al. [19] pursue a similar strategy but assume that the true camera pair can be approximated by cameras with fixed partial calibration (principal point in the image center and square pixels) and with unknown but constant focal length and estimate the fundamental matrix from the corresponding keypoints of two Local Affine Frames (LAFs) [18].

The calibration assumptions eliminate the need to construct artificial correspondences (dependent correspondences that are constructed from extracted features) as done in BEEM, but problems are expected on stereo pairs with zooming. Common to both methods is the close grouping of keypoints extracted from the distinguished region, which causes data-conditioning problems for model estimation.

### A. Contribution

This paper introduces two novel two-correspondence methods for estimating the fundamental matrix and extensively compares their performance with the methods listed in Table I (and in the above introductory text) on a standard test set of stereo pairs. Both proposed methods estimate the affine-fundamental matrix for RANSAC hypothesis generation but use different geometric primitives derived from detected image features in correspondence. The affine-fundamental matrix is an approximation to the global model and uses only reliable constraints from the local measurements. This is in contrast to previous methods, such as BEEM [10], where an approximation to the local image mapping is first used to generate additional point correspondences. The new point correspondences are then used as constraints to estimate the full model of epipolar geometry.

We show that on a large subset of the stereo pairs of the test set, using the affine-fundamental matrix for hypothesis generation gives performance comparable to the 7-point method when a local optimization step is used in RANSAC, and that it consistently performs better than hypothesizing the fundamental matrix with artificially constructed points as in BEEM or the essential matrix with unknown focal length from two LAFs as in Perdoch et al. [19].

All hypothesis generators are tested in the general model-optimization method introduced in Chum et al. [7], which provides a framework for solving a class of problems where the local optimization (LO) step of RANSAC includes estimating a more complex model (in terms of degrees of freedom).

### B. Structure of the paper

The rest of the paper is organized as follows: First, image features and geometric constructions used in this paper are reviewed in Section II. The analysis of the approximate affine epipolar geometry model generation is provided in Section III. The details of the local optimization are given in Section IV. Section V discusses the experimental evaluation of the methods; conclusions are drawn in Section VI.

hypothesis	denoted	relation	# corr	# models	primitive
7-point [11]	$F^7$	F	7	1-3	point
$F_A$ MLE [11]	$F_A^{LL}$	$F_A$	2	1	LAF
$F_A$ Arandjelović [1]	$F_A^{\circ\circ}$	$F_A$	2	0-4	ellipse
Perdoch [19]	$\{E, f\}^{LL}$	$E, f$	2	1-8	LAF
BEEM [10]	$F^{\circ\circ}$	F	2	1	oriented ell.

TABLE I: Tested RANSAC hypothesis generators. “relation” denotes what is estimated, “# corr” the number of corresponding primitives needed for estimation, “# models” the range of the number of relations generated for an estimate from one minimal sample, and “primitive” is the geometry extracted from detected image features that is needed to estimate the relation. Here  $E, f$  are used to denote the essential matrix and focal length.

## II. IMAGE FEATURES

In each image, Maximally Stable Extremal Regions (MSER) [16] are detected. An ellipse is fitted to each extremal region in an affine-covariant manner, using region moments of up to second order. The intensity patch covered by the ellipse is transformed to the unit circle and the dominant orientation is detected as in [14]. The SIFT descriptor [14] is extracted from each of the canonical patches.

Various methods of affine-covariant constructions of local coordinate systems called local affine frames (LAFs) were proposed by Obržálek and Matas [18]. Each of the LAFs is essentially defined by a triplet of points: its origin and two-extents. In the paper we have used the one of the simpler constructions, consisting of the center of gravity of the ellipse, the point on the ellipse coincident with the direction of the dominant orientation, and the point on the ellipse given by the direction perpendicular to the dominant orientation in the canonical pose.

## III. $F_A$ FOR HYPOTHESIS GENERATION

In the following paragraph we show that the first order Taylor approximation of the full epipolar constraint has a form of an affine epipolar constraint. Let  $F$  be a fundamental matrix and  $F_A$  an affine fundamental matrix

$$F = \begin{pmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{pmatrix}, \quad F_A = \begin{pmatrix} 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ a_3 & a_4 & a_5 \end{pmatrix}.$$

Let  $\mathbf{x} = (x, y, 1)^\top$ ,  $\mathbf{x}_0 = (x_0, y_0, 1)^\top$  be points in the first image and  $\mathbf{x}' = (x', y', 1)^\top$ ,  $\mathbf{x}'_0 = (x'_0, y'_0, 1)^\top$  be corresponding points in the second image. The first order approximation  $\hat{f}$  of the projective epipolar constraint  $\mathbf{x}'^\top F \mathbf{x}$  at a correspondence  $(\mathbf{x}_0, \mathbf{x}'_0)$  is

$$\mathbf{x}'^\top F \mathbf{x} \approx \mathbf{x}'_0^\top F \mathbf{x}_0 + J((x, y, x', y') - (x_0, y_0, x'_0, y'_0))^\top, \quad (1)$$

where  $J$  is the Jacobian of the projective epipolar constraint. Since  $\mathbf{x}'^\top F_A \mathbf{x} = a_1 x' + a_2 y' + a_3 x + a_4 y + a_5$  is a linear function in the image coordinates, the first order Taylor approximation

$f'$  of the projective epipolar constraint can be written as an affine epipolar constraint with the following entries of  $F_A$ :

$$\begin{aligned} a_1 &= f_1 x_0 + f_2 y_0 + f_3, & a_2 &= f_4 x_0 + f_5 y_0 + f_6 \\ a_3 &= f_1 x'_0 + f_4 y'_0 + f_7, & a_4 &= f_2 x'_0 + f_5 y'_0 + f_8, \\ a_5 &= \mathbf{x}'_0^\top F \mathbf{x}_0 - J(x_0, y_0, x'_0, y'_0)^\top. \end{aligned} \quad (2)$$

Thus, local to the correspondences used to estimate  $F_A$ , we expect  $F_A$  to also satisfy the epipolar constraint for projective cameras. This observation motivates the choice for using the affine fundamental matrix as an approximate model.

### A. $F_A$ for RANSAC hypothesis generation

The maximum likelihood estimate of  $F_A$  can be calculated from  $n \geq 4$  point correspondences [11]. We construct a two-correspondence  $F_A$  hypothesis generator by using the triples of points of two LAF correspondences. We use all point correspondences given by the two LAF correspondences which gives an over-determined estimate of  $F_A$ . We call this hypothesis generator “ $F_A$  MLE” and denote it  $F_A^{LL}$ .

For use in spatial verification in image retrieval engines, Arandjelović and Zisserman [1] derive a novel  $F_A$  estimator based on the parameterization of the affine-fundamental matrix introduced by Mendonça and Cipolla [17] that requires only two ellipse correspondences. An advantage of their parameterization is that orientations of the ellipses are not needed so dominant orientation estimation can be skipped during feature detection giving a speed-up. We call this hypothesis generator “ $F_A$  Arandjelović” and denote it  $F_A^{\circ\circ}$ .

## IV. GETTING TO F

Chum et al. [7] extend RANSAC by adding a generalized model optimization step that samples the consensus set to obtain higher accuracy estimates and, if needed, estimates a more complex model (a higher parameter model relative to the hypothesis generator). The approach is applicable to the approximate model setting: the point correspondences extracted from SIFT or LAF constructions are close to each other relative to the size of the image, which adversely affects data conditioning for model estimation, and the hypothesized model is, with the exception of BEEM, a lower parameter model than is required to model the scene geometry. RANSAC requires high accuracy estimation; otherwise, the required number of samples will greatly exceed what is predicted by theory [6]. Since speed is the primary motivation for using approximate models, local optimization (LO) is critical. RANSAC, as adapted from [7], is listed in Algorithm 1. The only component in Algorithm 1 that is specific to each approximate model is hypothesis generation. The remaining components are the same, and results from experiments are reported as such.

### A. Model Quality, Local Optimization and Stopping Criterion

Since LO design is known to be difficult, we use a proven estimator developed by Lebeda et al. [13] for the presented framework. Unlike the approximate models used for hypothesis generation, the LO step correctly models the geometry (modulo lens distortion), gives a maximum likelihood estimate of the fundamental matrix  $F$ , and an inlier-outlier labeling

**Input:**  $\mathcal{U}, k_{\max}$   
**Output:**  $F^*$   
 $k \leftarrow 0, \mathcal{J}_\theta \leftarrow \infty, \mathcal{J}_F \leftarrow \infty, \mathcal{I}^* \leftarrow \emptyset$   
**while**  $k < k_{\max}$  **do**  
  Make  $\mathcal{S}$  by sampling  $s$  correspondences from  $\mathcal{U}$   
  Estimate approximate hypothesis  $\theta$  consistent with  $\mathcal{S}$   
  Calculate  $\mathcal{J}_\theta$  from  $\theta$  as in eqn. 3  
  **if**  $\mathcal{J}_\theta < \mathcal{J}_\theta^*$  **then**  
     $\mathcal{J}_\theta^* \leftarrow \mathcal{J}_\theta$   
     $(F, \mathcal{J}_F) \leftarrow$  run Local Optimization  
    **if**  $\mathcal{J}_F < \mathcal{J}_F^*$  **then**  
       $F^* \leftarrow F$   
      Set  $\mathcal{I}^*$  as in eqn. 5  
      Update  $k_{\max}$  from eqn. 6 using  $\epsilon = \mathcal{I}^*/|\mathcal{U}|$   
  **end**  
**end**  
   $k \leftarrow k + 1$   
**end**  
**return**  $F^*$

**Algorithm 1:** RANSAC for approximate models. The set of tentative correspondences is denoted  $\mathcal{U}$ .

of tentative correspondences. The quality of the approximate model and fundamental matrix gotten from local optimization, denoted  $\mathcal{J}_\theta$  and  $\mathcal{J}_F$  respectively, are each measured by the robust cost function

$$\mathcal{J} = \sum_i \rho(e_i^2/(c\sigma^2)), \quad (3)$$

where  $\rho()$  is

$$\rho(e_i^2/(c\sigma^2)) = \begin{cases} e_i^2/(c\sigma^2) & \text{if } e_i^2/(c\sigma^2) < 1 \\ 1 & \text{otherwise} \end{cases}, \quad (4)$$

$e_i$  is the Sampson error of a keypoint computed from the given model,  $\sigma$  specifies feature localization noise and is set as in [13], and  $c$  is a tuning parameter (fixed for all experiments) which is set independently for each of  $\mathcal{J}_\theta, \mathcal{J}_F$  to account for the bias introduced by the approximate model. The number of inliers of  $F$  is given by

$$\mathcal{I} = |\{e_i \mid e_i^2/(c\sigma^2) < 1\}| \quad (5)$$

The necessary number of RANSAC iterations is estimated in the classical manner [9] using the best consensus set estimate  $\mathcal{I}^*$  after estimating  $F$  by LO,

$$k_{\max} = \frac{\log(1 - \eta_0)}{\log(1 - \epsilon^s)}, \quad (6)$$

where  $\epsilon = \mathcal{I}^*/|\mathcal{U}|$  is the proportion of inliers, and  $\eta_0$  is the confidence level (set to 0.99).

## V. EXPERIMENTS

The performance of BEEM, Perdoch, and the two proposed affine-fundamental matrix hypothesis generators were evaluated on a collection of 16 image pairs (Figure 1) for epipolar geometry estimation. These images were collated from test sets across a number of publications [2, 3, 4, 8, 15, 16, 19, 20, 22, 21] and are known to challenge state-of-the-art epipolar geometry estimators. Tentative correspondences were obtained by matching SIFT descriptors of MSER detections [14, 16].

	$F_A^{L\perp}$	$F_A^{\circ\circ}$	$\{E, f\}^{L\perp}$	$F^{\circ\circ}$
$F^7$	55.56%	59.59%	57.81%	57.40%
$F_A^{L\perp}$		53.45%	52.20%	52.84%
$F_A^{\circ\circ}$			48.74%	49.27%
$\{E, f\}^{L\perp}$				50.88%

TABLE II: Comparison of the RMS Sampson error. Entry  $(i, j)$  is the probability (expressed as a percentage) that the RMS Sampson error of an annotated correspondence set is less for epipolar geometry estimated by method  $i$  than  $j$  on the test set kusvod2.

The original BEEM method uses Difference of Gaussian (DoG) features [14] that provide localization, scale, and orientation only, thus the local image-to-image mapping is approximated by a similarity. For fairness of comparison, we extend the construction of the four-tuple of points to affine-covariant regions, so that the same geometric features are used in all compared methods. This construction approximates the local image-to-image mapping by an affine transformation, which is significantly more accurate in the wide-baseline setting, thus improving performance over standard BEEM construction with DoG features.

In all experiments, in each execution, the compared methods were forced to draw the same sequence of correspondence pairs to make the comparison less sensitive to the random nature of the random sampling.

### A. What is measured?

The accuracy of model estimation is given by two measures: the cardinality of the consensus correspondences (*i.e.* inlying feature correspondences to the model), which has been shown to be a good measure of model quality; and the root-mean-squared (RMS) distance of manually-annotated ground-truth points to their corresponding epipolar lines given by the estimated fundamental matrix. Precision is measured by the stability of the consensus set and variance of reprojection error for repeated model estimations.

In each of the methods, the construction of affine frames requires extra computation, and an affine frame can have multiple candidate matches because there can be several strong gradients in the descriptor region that can be used to fix the rotation of the LAF or SIFT affine frame construction. Furthermore, affine frames represent a small part of the image, making each of the above methods very sensitive to noise. Consequently, all prior methods require a local-optimization step [6] in the RANSAC loop. In contrast to the previous work, we seek a minimal solution that imposes no constraints on camera intrinsics—that is tolerant to noise without the need for local-optimization—and defines the state-of-the-art in both run-time performance and estimation stability.

### B. Results

Figure 2 plots the cumulative distribution of the RMS Sampson error on hand-annotated point correspondences collected over all executions of all stereo pairs. The plot shows that, over all, the highest accuracy was achieved by the 7-point algorithm, closely followed by the  $F_A^{L\perp}$  method. The other three



Fig. 1: The kusvod2 test set used for evaluating fundamental matrix estimators. Each stereo pair has manually annotated ground truth correspondences that are used to assess the accuracy of the estimated epipolar geometry. The test set with annotated point correspondences is available at <http://cmp.felk.cvut.cz/data/geometry2view/>.

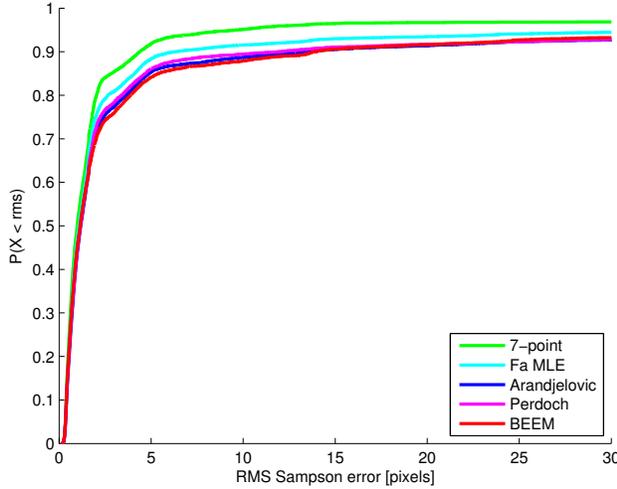


Fig. 2: Empirical cumulative density function of the RMS Sampson error of  $F$  as estimated by the tested methods. Error is calculated on hand-annotated point correspondences for each stereo pair in the test set.

	$F_A^{L,L}$	$F_A^{O,O}$	$\{E, f\}^{L,L}$	$F^{O,O}$
$F^7$	0.24%	1.14%	1.56%	4.24%
$F_A^{L,L}$		50.02%	62.75%	76.62%
$F_A^{O,O}$			52.40%	68.03%
$\{E, f\}^{L,L}$				60.84%

TABLE III: Comparison of RANSAC samples drawn. Entry  $(i, j)$  is the probability (expressed as a percentage) that the number of samples required for RANSAC termination is fewer for method  $i$  than  $j$  on the test set kusvod2.

methods, in terms of accuracy, give similar results to each other but perform worse than the first two methods. They rank descending as  $\{E, f\}^{L,L}$ ,  $F_A^{O,O}$  and BEEM.

For the average number of models per sample, important for assessing the cost of verifying hypotheses, each of  $F_A^{L,L}$  and BEEM generate a single model per sample,  $\{E, f\}^{L,L}$  method

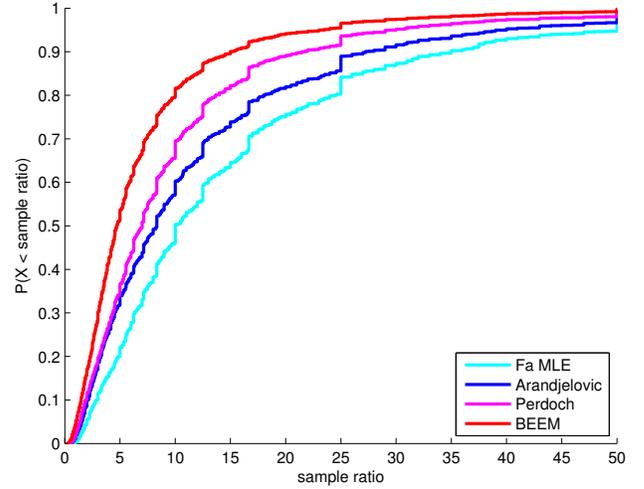


Fig. 3: Empirical cumulative density function of the sample ratio, calculated as the ratio of the number of RANSAC samples required by the 7-point method to the denoted approximate model.

generates 1.2 to 1.9 models per sample in practice, the 7-point algorithm generates 2.38 models per sample (the value is taken from [5], as the executable from Lebeda et al. [13] does not provide the number of models), and  $F_A^{O,O}$  generates 3.0 to 3.4 models per sample, which suggests that it is the most demanding of the methods in terms of hypothesis verification runtime.

Figure 3 shows the relative speed-up to the baseline seven-point algorithm measured by the number of samples. The lower the curve – the lower fraction of execution with speed-up lower than a given speed-up – the higher relative speed up. The highest speed up is achieved by the  $F_A^{L,L}$ . Since this method also generates the lowest number of models per sample, and the model estimate is fast, it is the fastest of the hypothesis generators. The advantage of  $F_A^{O,O}$  method is, that unlike the other methods, it does not require point correspondences or dominant orientation for model estimation, so some speed may be gained during feature extraction.

Tables II and III directly compare pairs of methods by accuracy of fundamental matrix estimation and by the number of samples required for RANSAC termination respectively. The entries show the probability that a method listed by row will outperform a method listed by column on the kusvod2 data set. Detailed results on a subset of the kusvod2 dataset are shown in Table IV. Denotations are defined as follows:  $I$ , absolute number, standard deviation, and minimal and maximal number of inliers detected;  $I(\%)$ , percentage of inliers detected from the set of tentative correspondences;  $Samp$ , number of samples drawn by RANSAC with the specified hypothesis generator;  $Models$ , mean number of models verified per sample;  $Error$ , average Sampson error on manually annotated ground truth point correspondences (they are not used in the estimation process);  $LO\ count$ , number of local optimization executions.

The  $Inlss$  and  $HInlss$  histograms embedded in Table IV provide a concise overview of the stability of epipolar geometry estimation over 500 repeated executions and are computed as follows:  $Inlss$  is the probability of a particular correspondence being an inlier, where correspondence ids on the horizontal axis are ordered so that the values on the vertical axis, the fraction of executions the correspondence has been labeled as an inlier, are non-increasing.  $HInlss$  is the histogram of the probability of a correspondence being an inlier, where the horizontal axis is the percentage of total executions (out of 500) and the vertical axis is the number of tentative correspondences that have been labeled as inliers.

## VI. CONCLUSION

This paper proposes two two-correspondence methods for estimating the fundamental matrix. Each of the proposed methods uses the affine-fundamental matrix as an approximate model hypothesis generator for RANSAC. The affine-fundamental matrix is a local approximation to the epipolar geometry for perspective cameras, and it uses only reliable constraints from local measurements. The full model is estimated in a generalized model optimization framework similar to [7, 13].

Experimental validation shows that using the approximation of the global model from two LAFs outperforms previously proposed approximate models which include a technique that uses an approximation of the local image mapping to construct artificial additional point correspondences for full model estimation as well as an estimator that adds algebraic constraints by enforcing certain calibration parameters on the approximate model [10, 19]. On a large subset of the tested stereo pairs, using the affine-fundamental matrix for hypothesis generation gives fundamental matrix estimation accuracy comparable to the 7-point method as measured by root mean square Sampson error on manually annotated correspondences while achieving a significant speed-up in terms of RANSAC samples required, number of models verified, and number of local optimization executions made. On nearly the entire test set  $F_A$  MLE ( $F_A^{MLE}$ ) was more accurate than all other approximate hypothesis generators and was at least as fast, so we recommend it in lieu of the alternatives.

## ACKNOWLEDGMENT

James Pritts and Ondřej Chum were supported by GACR P103/12/2310 and MSMT LL1303 ERC-CZ grants and Jiří

Matas by EC project FP7-ICT-288587 MASELTOV. The authors thank Zuzana Kukulova for providing the 6-pt relative pose solver [12].

## REFERENCES

- [1] R. Arandjelović and A. Zisserman. Efficient image retrieval for 3D structures. In *Proc. of BMVC*, 2010.
- [2] J. Cech, J. Matas, and M. Perdoch. Efficient sequential correspondence selection by cosegmentation. *Pattern Analysis and Machine Intelligence*, 32(9):1568–1581, 2009.
- [3] S. Choi, T. Kim, and W. Yu. Performance evaluation of RANSAC family. In *Proc. of BMVC*, 2009.
- [4] O. Chum and J. Matas. Matching with PROSAC – progressive sample consensus. In *Proc. of CVPR*, pages 220–226, 2005.
- [5] O. Chum and J. Matas. The geometric error for homographies. *Pattern Analysis and Machine Intelligence*, 2008.
- [6] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. In *DAGM Symposium*, 2003.
- [7] O. Chum, J. Matas, and Š. Obdržálek. Enhancing RANSAC by generalized model optimization. In *Proc. of ACCV*, volume 2, pages 812–817, 2004.
- [8] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *Proc. of CVPR*, pages 772–779, 2005.
- [9] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *CACM*, 24(6):381–395, June 1981.
- [10] L. Goshen and I. Shimshoni. Balanced exploration and exploitation model search for efficient epipolar geometry estimation. *Pattern Analysis and Machine Intelligence*, 30(7):1230–1242, 2008.
- [11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Second edition, 2004.
- [12] Z. Kukulova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *Proc. BMVC*, pages 56.1–56.10, 2008.
- [13] K. Lebeda, J. Matas, and O. Chum. Fixing the locally optimized ransac. In *Proc. of BMVC*, pages 1–11, 2012.
- [14] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [15] D. Martinec and T. Pajdla. 3D reconstruction by fitting low-rank matrices with missing data. In *Proc. of CVPR*, pages 198–205, 2005.
- [16] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. of BMVC*, pages 384–396, 2002.
- [17] P. Mendonça and R. Cipolla. Estimation of epipolar geometry from apparent contours: Affine and circular motion cases. In *Proc. of CVPR*, 1999.
- [18] Š. Obdržálek and J. Matas. Object recognition using local affine frames on distinguished regions. In *Proc. of BMVC*, pages 113–122, 2002.
- [19] M. Perdoch, J. Matas, and O. Chum. Epipolar geometry from two correspondences. In *Proc. of ICPR*, pages 215–220, 2006.
- [20] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool. Automated reconstruction of 3D scenes from sequences of images. *ISPRS Journal Of Photogrammetry And Remote Sensing*, 55(4): 251–267, 2000.
- [21] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proc. of BMVC*, pages 412–422, 2000.
- [22] G. Yang, C. Stewart, M. Sofka, and C. Tsai. Registration of challenging image pairs: Initialization, estimation, and decision. *Pattern Analysis and Machine Intelligence*, 29(11):1973–1989, 2007.

Solver → Detectors → Descriptors →		$F^r$ MSER+ MSER- SIFT	$F_u^+$ MSER+ MSER- SIFT	$F_u^0$ MSER+ MSER- SIFT	$F_u^-$ MSER+ MSER- SIFT	$\{E_r, f\}^{1-\alpha}$ MSER+ MSER- SIFT	$F^{00}$ MSER+ MSER- SIFT	
Image		500 runs, $\sigma = 1$ , conf = 99%						
Qtr ↓		500 runs, $\sigma = 1$ , conf = 99%						
booksh	I	80.8 ±1.5 (72-84)	80.2 ±2.1 (72-84)	80.5 ±2.0 (72-84)	80.5 ±2.0 (72-84)	80.5 ±1.6 (72-84)	79.6 ±3.0 (49-84)	
	I (%)	73.4 ±1.3 (65-76)	72.9 ±1.9 (65-76)	73.1 ±1.8 (66-76)	73.1 ±1.8 (66-76)	73.2 ±1.4 (65-76)	72.3 ±2.7 (45-76)	
	Snapp	45.8 ±6.3 (24-57)	2.3 ±1.4 (1-4)	2.1 ±1.3 (1-7)	2.1 ±1.3 (1-7)	3.5 ±2.3 (1-14)	5.8 ±4.0 (1-27)	
corr	Mean Models	N/A	1.0	3.3	3.3	1.5	1.0	
	Error	1.49 ±1.70(0.5-15.4)	2.24 ±3.20 (0.5-14.1)	1.91 ±2.75 (0.4-14.1)	1.91 ±2.75 (0.4-14.1)	1.56 ±1.62 (0.5-14.3)	3.04 ±4.16 (0.5-20.9)	
	LO count	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	
Kyoto	I	171.0 ±2.0 (165-176)	170.5 ±2.3 (163-176)	170.4 ±2.3 (161-178)	170.4 ±2.3 (161-178)	170.5 ±2.2 (164-177)	170.3 ±2.3 (163-176)	
	I (%)	78.5 ±0.9 (76-81)	78.2 ±1.1 (75-81)	78.2 ±1.1 (74-82)	78.2 ±1.1 (74-82)	78.2 ±1.0 (75-81)	78.1 ±1.1 (75-81)	
	Snapp	29.7 ±6.8 (15-49)	1.7 ±0.9 (1-6)	1.9 ±1.2 (1-7)	1.9 ±1.2 (1-7)	2.6 ±1.7 (1-10)	3.8 ±2.4 (1-12)	
plant	Mean Models	N/A	1.0	3.0	3.0	1.9	1.0	
	Error	0.47 ±0.13 (0.2-0.9)	0.51 ±0.14 (0.2-1.0)	0.51 ±0.14 (0.2-1.1)	0.51 ±0.14 (0.2-1.1)	0.51 ±0.13 (0.1-1.2)	0.53 ±0.14 (0.2-1.2)	
	LO count	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	
valbonne	I	988.4 ±2.8 (980-996)	988.3 ±2.9 (982-997)	988.4 ±3.0 (981-998)	988.4 ±3.0 (981-998)	988.4 ±2.9 (979-998)	988.6 ±3.0 (981-999)	
	I (%)	79.8 ±0.2 (79-80)	79.8 ±0.2 (79-80)	79.8 ±0.2 (79-81)	79.8 ±0.2 (79-81)	79.8 ±0.2 (79-81)	79.8 ±0.2 (79-81)	
	Snapp	23.9 ±8.0 (13-50)	2.7 ±1.7 (1-9)	4.4 ±2.8 (1-17)	4.4 ±2.8 (1-17)	5.4 ±3.2 (1-21)	6.2 ±3.7 (1-23)	
wall	Mean Models	N/A	1.0	3.3	3.3	1.5	1.0	
	Error	1.07 ±0.36 (0.5-2.5)	1.05 ±0.32 (0.5-2.4)	1.13 ±0.38 (0.4-2.7)	1.13 ±0.38 (0.4-2.7)	1.10 ±0.36 (0.4-2.5)	1.09 ±0.36 (0.4-2.4)	
	LO count	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	
wash	I	43.5 ±0.7 (40-45)	43.2 ±1.6 (17-45)	42.4 ±3.3 (16-45)	42.4 ±3.3 (16-45)	39.8 ±7.7 (14-45)	37.2 ±8.8 (13-45)	
	I (%)	63.9 ±1.0 (59-66)	63.6 ±2.4 (25-66)	62.4 ±4.9 (24-66)	62.4 ±4.9 (24-66)	58.5 ±11.3 (21-66)	54.8 ±12.9 (19-66)	
	Snapp	69.7 ±8.3 (53-128)	3.5 ±2.2 (1-12)	6.6 ±3.0 (1-20)	6.6 ±3.0 (1-20)	8.1 ±2.8 (1-17)	15.1 ±8.2 (1-47)	
zoom	Mean Models	N/A	1.0	3.4	3.4	1.2	1.0	
	Error	1.68 ±1.25(0.9-20.9)	1.85 ±2.18 (0.9-36.8)	3.45 ±5.78 (0.9-43.2)	3.45 ±5.78 (0.9-43.2)	10.28±23.00(0.9-161.1)	10.28±13.11 (0.9-63.2)	
	LO count	1.3 ±0.5 (1-4)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.2 (0-1)	1.0 ±0.0 (1-1)	
booksh	I	65.0 ±0.7 (60-66)	61.5 ±5.7 (35-66)	61.8 ±5.7 (39-67)	61.8 ±5.7 (39-67)	61.9 ±6.2 (33-66)	63.3 ±4.6 (23-66)	
	I (%)	73.8 ±0.8 (68-75)	69.9 ±6.5 (40-75)	70.2 ±6.5 (44-76)	70.2 ±6.5 (44-76)	70.3 ±7.1 (38-75)	72.0 ±5.3 (26-75)	
	Snapp	45.7 ±7.1 (21-54)	4.8 ±2.1 (1-16)	4.4 ±2.0 (1-12)	4.4 ±2.0 (1-12)	6.5 ±2.8 (1-16)	8.1 ±4.4 (1-27)	
wall	Mean Models	N/A	1.0	3.0	3.0	1.2	1.0	
	Error	6.46 ±4.38(1.0-43.2)	26.32±35.75(1.0-138.9)	26.71±38.91(1.0-156.8)	26.71±38.91(1.0-156.8)	14.74±23.41(0.9-153.3)	14.74±23.41(0.9-153.3)	
	LO count	1.0 ±0.0 (1-2)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	
wash	I	197.2 ±0.4 (196-199)	195.5 ±9.2 (109-199)	193.4 ±13.5 (109-199)	193.4 ±13.5 (109-199)	194.9 ±10.2 (113-199)	191.4 ±19.1 (88-199)	
	I (%)	88.4 ±0.2 (88-89)	87.7 ±4.1 (49-89)	86.7 ±6.1 (49-89)	86.7 ±6.1 (49-89)	87.4 ±4.6 (51-89)	85.8 ±8.6 (39-89)	
	Snapp	17.0 ±8.3 (6-50)	3.4 ±1.7 (1-11)	4.9 ±2.3 (1-13)	4.9 ±2.3 (1-13)	4.3 ±2.5 (1-15)	8.8 ±5.6 (1-33)	
zoom	Mean Models	N/A	1.0	3.2	3.2	1.5	1.0	
	Error	0.56 ±0.14 (0.2-0.9)	1.31 ±5.31 (0.3-66.3)	1.90 ±5.90 (0.3-68.1)	1.90 ±5.90 (0.3-68.1)	1.30 ±4.09 (0.3-51.0)	4.07 ±17.62(0.3-240.4)	
	LO count	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	
booksh	I	124.8 ±0.9 (122-126)	118.5 ±16.5 (57-126)	106.1 ±25.4 (55-126)	106.1 ±25.4 (55-126)	118.3 ±17.5 (30-126)	119.4 ±15.7 (57-126)	
	I (%)	86.0 ±0.6 (84-87)	81.7 ±11.4 (39-87)	73.2 ±17.5 (38-87)	73.2 ±17.5 (38-87)	81.6 ±12.0 (21-87)	82.4 ±10.9 (39-87)	
	Snapp	21.9 ±9.5 (7-50)	2.5 ±1.4 (1-8)	6.0 ±3.2 (1-21)	6.0 ±3.2 (1-21)	6.5 ±3.4 (1-20)	5.2 ±2.9 (1-16)	
zoom	Mean Models	N/A	1.0	3.3	3.3	1.1	1.0	
	Error	0.55 ±0.15 (0.3-1.0)	2.20 ±6.83 (0.3-79.6)	8.84 ±20.15(0.3-133.5)	8.84 ±20.15(0.3-133.5)	3.21 ±12.12(0.2-126.7)	2.39 ±8.13 (0.3-75.7)	
	LO count	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (0-1)	1.0 ±0.0 (1-1)	
booksh	I	113.9 ±1.1 (108-117)	112.4 ±6.3 (65-117)	111.9 ±7.4 (67-117)	111.9 ±7.4 (67-117)	102.3 ±17.3 (54-117)	110.0 ±10.4 (65-117)	
	I (%)	59.3 ±0.6 (56-61)	58.5 ±3.3 (34-61)	58.3 ±3.8 (35-61)	58.3 ±3.8 (35-61)	57.3 ±9.0 (28-61)	57.3 ±5.4 (34-61)	
	Snapp	116.2 ±9.8 (96-181)	5.4 ±2.9 (1-18)	6.1 ±2.8 (1-17)	6.1 ±2.8 (1-17)	8.2 ±3.2 (1-20)	12.5 ±6.3 (1-35)	
zoom	Mean Models	N/A	1.0	3.3	3.3	1.6	1.0	
	Error	0.71 ±0.28 (0.4-3.5)	1.51 ±6.83 (0.2-143.7)	1.75 ±4.59 (0.3-39.7)	1.75 ±4.59 (0.3-39.7)	11.24±23.34(0.4-150.5)	3.43 ±11.50(0.4-125.3)	
	LO count	1.8 ±0.8 (1-5)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	1.0 ±0.0 (1-1)	

TABLE IV: Detailed results on selected image pairs.