ENHANCING RANSAC BY GENERALIZED MODEL OPTIMIZATION

*Ondřej Chum, Jiří Matas and Štěpán Obdržálek**

Center for Machine Perception, Dpt. of Cybernetics, Faculty of Elect. Engineering,
Czech Technical University, 121 35 Prague, Karlovo nam. 13, Czech Republic
[chum,matas,xobdrzal]@cmp.felk.cvut.cz

ABSTRACT

An extension of the RANSAC procedure is proposed. By adding a generalized model optimization step (the LO step) applied only to models with a score (quality) better than all previous ones, an algorithm with the following desirable properties is obtained: a near perfect agreement with theoretical (i.e. optimal) performance and lower sensitivity to noise and poor conditioning. The chosen scheduling strategy is shown to guarantee that the optimization step is applied so rarely that it has minimal impact on the execution time.

The potential of the LO step is demonstrated on two new matching algorithms that are straightforward implementation of LO-RANSAC: 1. an algorithm for simultaneous estimation of epipolar geometry and radial distortion and 2. an algorithm estimating epipolar geometry from three region-to-region correspondences.

Experiments show that the new estimators have performance superior to the state-of-the-art. The latter algorithm speeds up the estimation of epipolar geometry more than thousand times and thus permits solving complex problems with inlier ratio below 10%. The former represents a class of algorithms where the LO step includes switching to a more complex model (with more degrees of freedom). Surprisingly, due to the increased number of inliers of the more complex (and more precise) model, the LO step significantly reduces the execution time.

1. INTRODUCTION

The RANSAC algorithm [2, 11, 10] is a robust estimator widely used in the field of computer vision, especially in the two view geometry – either homography or epipolar geometry (EG) – estimation [5, 9, 13, 7, 3, 8, 6].

The main contribution of the paper is an extension of the RANSAC algorithm. By adding a generalized model optimization step (the LO step) applied only to models with a score (quality) better than all previous ones, an algorithm with the following desirable properties is obtained: a near perfect agreement with theoretical performance and lower sensitivity to noise and poor conditioning.

Fig. 1: Correspondences established by the proposed algorithm on the 'Flower' pair. Tentative correspondences contained only 7% of inliers. Some of the corresponding points are connected between the images to visualize the correspondence.

The potential of the proposed extension of the RANSAC algorithm is supported by two new matching algorithms exploiting LO-RANSAC: First, a modified algorithm for simultaneous estimation of epipolar geometry (EG) and radial distortion (RD) [3] is described. Experimental validation shows that the new estimator is superior to known algorithms in quality (the number of detected matches), precision and speed of the matching. Second, an algorithm estimating epipolar geometry from three region-to-region cor-

*This work was supported by the following projects: IST-2001-32184, CTU0306013, CTU0307013, MŠMT LN00B096.

respondences is introduced. Exploiting the affine-invariant local frames described in [7], three point-to-point correspondences are found for each region-to-region correspondence. The expected run-time then falls from $\mathcal{O}(\varepsilon^{-7})$ to $\mathcal{O}(\varepsilon^{-3})$. The straightforward consequence is a significant enlargement of the class of problems that are solvable[1].

The rest of the paper is structured as follows: The LO-RANSAC algorithm [1] is reviewed in Section 2. Algorithm for estimation of EG under a more complex camera model which includes radial distortion is introduced and experimentally verified in Section 3. Details of the new algorithm that computes EG from three region-to-region correspondences are given in Section 4. The paper is concluded in Section 6.

2. LO-RANSAC

The structure of the RANSAC algorithm is simple but powerful. Repeatedly, subsets are randomly selected from the input data points (a set of points from $\mathbb{R}^2$ when a line is being estimated, a set of tentative correspondences when EG is being estimated) and model parameters fitting the sample are computed. The size m of the random samples is the smallest sufficient to determine model parameters. In a second step (verification), the quality of the model parameters is evaluated on the full data set. Different cost functions may be used [12] for the evaluation, the standard being the the cardinality of the support, i.e. the number of data points consistent with the model. Let P be the probability that an uncontaminated sample of size m is randomly selected from a set $\mathcal{U}$ of N data points is

$$P = \frac{\binom{I}{m}}{\binom{N}{m}} = \prod_{j=0}^{m-1} \frac{I-j}{N-j} \approx \varepsilon^m, \tag{1}$$

where ε is the fraction of inliers $\varepsilon = I/N$. The sampling process is terminated [2, 11] when the likelihood of finding a better model becomes low. Assuming (incorrectly, as we later show) that P is also the probability of finding a correct solution, the termination criterion is formulated as follows. The sampling is repeated until the probability η of missing a set of inliers $\mathcal{I}$ of size $I = |\mathcal{I}|$ within k samples falls under a predefined threshold η_0,

$$\eta = (1-P)^k. \tag{2}$$

The number of samples that has to be drawn to satisfy $\eta \leq \eta_0$ is

$$k_{\eta_0,\varepsilon} = \log(\eta_0)/\log(1-P). \tag{3}$$

[1]The idea of using multiple points in the estimation process is in principle simple. However, since the three points associated with a single region are in close proximity, the precision of the estimated epipolar geometry may be questioned. The experiments confirmed, that acquisition of a new local optimization step into the RANSAC algorithm was essential to solve the problem.

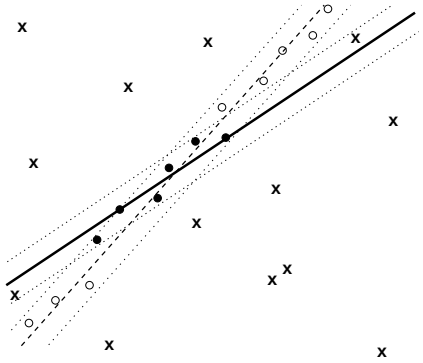

Fig. 2: An example of a β-sample (solid line) in line detection by RANSAC. The two point outlier-free sample led to a model that is not consistent with all inliers (circles). The dashed line represents the optimal model, the dotted lines depict the tolerance areas.

However, it has been observed experimentally [10, 1] that RANSAC often runs much longer than predicted by the mathematical model of the process, i.e. by equations (1) and (3). To explain the phenomenon, we introduce the following terminology:

A β**-sample** is a subset of the set of inliers $\mathcal{I}$. The set $\mathcal{B}$ of all β-samples of size m has $\binom{I}{m}$ elements.
An α**-sample** is a subset of the set of inliers $\mathcal{I}$ such that the model computed from the sample is consistent with all inliers. Both β-samples and α-samples are uncontaminated by definition, i.e. they are free of outliers.
A γ**-sample** is a subset of all tentative correspondences containing at least one incorrect correspondence (i.e. at least one outlier). This type of sample is often called 'contaminated'.

The discrepancy between the predicted and observed RANSAC run times can be explained in terms of cardinalities of the set $\mathcal{B}$ of β-samples and the set $\mathcal{A}$ of α-samples. The termination criterion tacitly assumes, that all β-samples are α-samples, i.e. that every model computed from *an uncontaminated sample* is consistent with *all* inliers. This favorable situation is rare in practice. Since RANSAC generates hypotheses from minimal sets, both noise in the data and poor conditioning of the model computation can 'remove' an uncontaminated sample from the set of α-samples $\mathcal{A}$. The situation is demonstrated in the context of line detection in Fig. 2. In [1], it is shown that in the process of estimating either homography or EG from point-to-point correspondences the difference in the $\mathcal{A}$ and $\mathcal{B}$ sets leads to an increase in the number of RANSAC cycles by a factor of two to three. In the case of epipolar geometry estimation from three region-to-region correspondences (analyzed in Section 4), the ratio of cardinalities of $\mathcal{A}$ and $\mathcal{B}$ is so low that it renders the approach, in connection with standard RANSAC, impractical.

The LO-RANSAC is based on the observation that virtually all models estimated from an uncontaminated min-

imal sample contain large fraction of inliers within their support. An optimization[2] process starting from the so-far-the-best hypothesized model is therefore inserted into RANSAC. *Applying the proposed optimization step produces an algorithm with a near perfect agreement with theoretical (i.e. optimal) performance*. In other words, LO-RANSAC makes the sets $\mathcal{A}$ and $\mathcal{B}$ almost identical. Therefore eq. (2) becomes valid for LO-RANSAC. The structure of the LO-RANSAC algorithm is presented in Alg. 1.

Repeat until the probability (2) of finding model with support larger than I^* in k-th step falls under predefined threshold η_0:

1. Select a random sample of minimum size m from $\mathcal{U}$.
2. Estimate model parameters consistent with the sample.
3. Calculate the support I_k of the model, i.e. the data points with error smaller than a predefined threshold θ. If $I_k > I_j$ for all $j < k$ (i.e. when a new maximum is reached), then run:
   LO step. Apply optimization. Store the best model found and its support I^* ($I^* \geq I_k$ due to the optimization).

Algorithm 1: The structure of LO-RANSAC. Note, that the algorithm is terminated based on the optimized support I^*, whereas execution of the LO step depends on supports of sampled hypotheses I_j.

2.1. The additional computational cost

The LO step is carried out only if a new maximum in the number of inliers is reached, i.e. when standard RANSAC stores its so-far-the-best result. How often does this happen? The number of data points consistent with a model from a randomly selected sample can be thought of as a random variable with an unknown density function. This density function is the same for all samples, so the probability that k-th sample will be the best so far is $1/k$. The average number of maxima reached within k samples is thus

$$\sum_{x=1}^{k} \frac{1}{x} \leq \int_{1}^{k} \frac{1}{x}\, dx + 1 = \log k + 1. \tag{4}$$

The logarithmic growth of the number of LO step invocations as a function of the number of hypothesize-and-verify cycles allows application of relatively computationally expensive optimization methods without an impact on the overall speed of the algorithm.

2.2. The LO step

Different methods of the best model optimization with respect to the two view geometry estimation were proposed and tested [1]. The following procedure performed the best.

[2]Note, that the LO-RANSAC does *not* try to compete with the bundle adjustment methods. The aim is to provide a better starting point for the bundle adjustment than standard RANSAC in shorter time.

Fig. 3: 'Orange house', 45% of inliers. A low radial distortion example.

Constant number (twenty in our experiments) of samples are drawn only from $\mathcal{I}_k$, while the verification is performed on the set of *all* data points $\mathcal{U}$ (so called 'inner' RANSAC). Since the proportion of inliers in $\mathcal{I}_k$ is high, there is no need for the size of sample to be minimal. The problem has shifted from minimizing the probability of including an outlier into the sample (the reason for choosing minimal sample size) to the problem of reduction of the influence of the noise on model parameters. The size of the sample is therefore selected to maximize the probability of drawing an α-sample. In a final step, model parameters are 'polished' by an iterative reweighted least squares technique.

In Section 3 we show that the LO-RANSAC algorithm can not only maximize the support of model by improving its precision, but it can also switch to a more complex model with more accurate fit.

3. ESTIMATING RADIAL DISTORTION

The benefits of exploiting a more complex model within the LO step can be demonstrated on the problem of simultaneous estimation of EG and radial distortion (RD). RD is a deviation from the pinhole model commonly encountered in cameras, especially with wide-angle lenses. Fitzgibbon [3] introduced an algorithm for joint estimation of EG and RD given 9 point correspondences. It follows from eq. (1) that the price paid for the inclusion of Fitzgibbon's RD estimation is an increase in the number of samples drawn by a factor of $1/\varepsilon^2$. Since typically $\varepsilon \in [0.1, 0.5]$, the nine point RANSAC for EG and RD estimation (9pt RANSAC) is 4 to 100 slower than the standard 7pt RANSAC.

3.1. The LO-RANSAC-RD algorithm

We present the simultaneous estimation of EG and RD as an algorithm in the LO-RANSAC framework. The algorithm draws 7 correspondence samples to estimate EG without RD in the hypothesize-verify loop and includes RD model in the LO step. Such sampling strategy ensures that LO-RANSAC-RD has the same time complexity $\mathcal{O}(\varepsilon^{-7})$ as the 7pt RANSAC. To parameterize the RD model, we have chosen the division model [3]

$$\mathbf{p} = \frac{1}{1 + \lambda |\mathbf{x}|^2} \mathbf{x},$$

Fig. 4: 'Courtyard QY' image pair, 48% of inliers. High radial distortion example. Central part (1400 x 1400 pixels, view field 100°) of the fish-eye images. Original first image on the left, the second (different) image with RD removed on right.

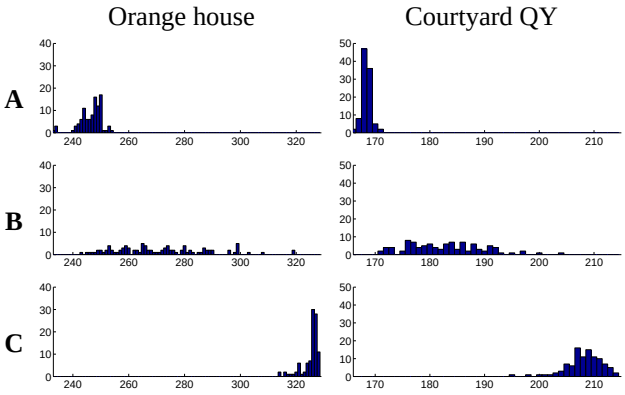

Fig. 5: Histograms of the numbers of inliers detected by different RANSAC algorithms: A - 7pt LO-RANSAC with no RD estimation, B - 9pt RANSAC, C - LO-RANSAC-RD.

where λ is the only model parameter, $\mathbf{x}$ stands for the measured point, $|\mathbf{x}|$ for the distance of $\mathbf{x}$ to the optical center (center of the image) and $\mathbf{p}$ for the undistorted point so that the epipolar constraint can be written as $\mathbf{p}'\mathbf{F}\mathbf{p} = 0$. EG and RD were simultaneously estimated in the LO step solving quadratic eigenvalue problem as in [3].

3.2. Experiments

Performance of three algorithms, 7pt LO-RANSAC (A), 9pt RANSAC (B) and LO-RANSAC-RD (C), was compared on image pairs with low RD (Orange house, Fig. 3) and high RD (Courtyard QY, Fig. 4). The number of detected inliers is shown in Fig. 5. Alg. B finds more inliers than A because it uses a more precise model. Alg. C finds more inliers than B due to the LO step. The speed of A,B and C is measured by the number of samples drawn (Tab. 1). Alg. B is the slowest, as its time complexity is $\mathcal{O}(\varepsilon^{-9})$, compared with $\mathcal{O}(\varepsilon^{-7})$ of A and C. As a consequence of eq. (3), C terminates much earlier than A since it finds a higher number of inliers. Finally, the stability of the radial distortion estimation was measured. The graphs of the distribution of estimated parameter λ, depicted in Fig. 6, show that C is more stable than B – the variation of λ is smaller.

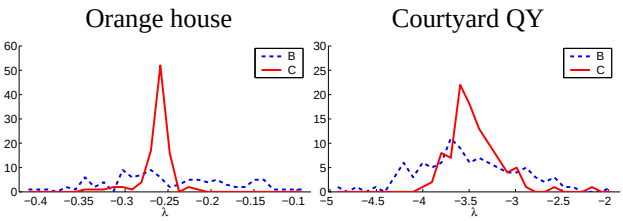

Fig. 6: The distribution of estimated parameter λ of radial distortion for 9pt RANSAC (B dashed) and LO-RANSAC-RD (C solid).

RANSAC	7pt LO	9pt	LO–RD
Orange house	5 528	31 456	790
Courtyard QY	1 861	6 863	432

Table 1: The average number of samples drawn over 100 runs of different RANSAC algorithm. Almost 40 times speed-up between 9pt RANSAC and LO-RANSAC-RD was observed.

4. EG FROM THREE CORRESPONDENCES

In wide-baseline matching, the process of selection of tentative correspondences often produces region-to-region mappings. However, the strong constraints the mappings provide are ignored in the EG estimation stage. The fundamental matrix is computed from seven point-to-point correspondences, with each region-to-region mapping providing just a single point-to-point correspondence [13, 9, 7] .

In this section, we assume that a set of region-to-region tentative matches is available, and that three independent point-to-point matches can be obtained per each. In experiments, we used the output of the method published in [7], where the triplet of points originating from one region is called 'local affine frame' (LAF).

4.1. The 3LAF LO-RANSAC algorithm

The algorithm is a straightforward application of LO-RANSAC. To hypothesize a model of epipolar geometry, random samples of three region correspondences are drawn. Three region correspondences give nine point correspondences. These are used to estimate the fundamental matrix F using the linear eight-point algorithm [4]. The LO step is applied (as always) only to the models that are so far the best and includes both the 'inner' RANSAC and iterative polishing, as described in Section 2. A region correspondence is consistent with a hypothesized epipolar geometry iff all three points are consistent.

4.2. Experiments

To highlight the advantage of 3LAF LO-RANSAC, tests were carried out on two image pairs (Fig. 1 and 8) with only about

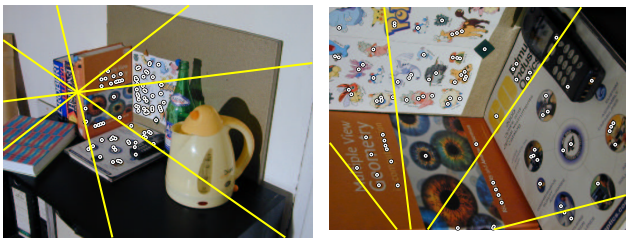

Fig. 7: Epipolar lines and 84 detected correspondences (markers) superimposed over the 'Bookshelf' pair.

Method	EG consist.	iterations
Flower, **518** tentative corr., **7**% inliers		
7pt LO-RANSAC	N/A	$\approx 684\,000\,000$
3LAF RANSAC	25	47 668
3LAF LO-RANSAC	36	14 880
Bookshelf, **201** tentative corr., **42**% inliers		
7pt LO-RANSAC	83	1705
3LAF RANSAC	47	245
3LAF LO-RANSAC	84	41
Ascona, **584** tentative corr., **20**% inliers		
7pt LO-RANSAC	116	284 868
3LAF RANSAC	65	2 867
3LAF LO-RANSAC	116	384

Table 2: Summary of experimental results. Number of correspondences found consistent with the epipolar geometry and the number of RANSAC iterations required to reach the solution. Note that all the numbers are random variables.

7% and 20% of tentative correspondences correct respectively. The bookshelf test pair (Fig. 7) represents an indoor scene with large scale difference between the two views.

Results of the conducted experiments are summarized in Tab. 2. For the 'Flower' pair, the fraction of inliers is so low that the standard seven-point method failed. In the other two pairs, a significant speed-up measured by the number iterations was achieved.

It is important to note that when applying the 3-frame RANSAC without the LO step (3LAF RANSAC), the set of detected inliers is significantly smaller, as shown in the middle column of Tab. 2. We believe that this is due to the fact that local affine frames are typically very small and the three points from a single region lie in near proximity. Consequently, the EG estimated from a minimal number of correspondences, as well as its support set, are unstable.

Clearly, application of the LO step is a very important ingredient of the newly proposed algorithm. As a final remark, we note that requiring all three point-to-point correspondence that form a region-to-region correspondence to obey the EG constraint also reduces the number of false positive matches, as the probability that a random correspondence will satisfy the epipolar constraint is decreased.

Fig. 9: The 'Lobby' pair, 28% of inliers. Central parts of the fish-eye images. Original first image on the left, the second image with radial distortion removed on right. Correspondences detected by 3LAF LO-RANSAC-RD superimposed over the images (markers).

5. ALL-IN-ONE.

Estimation of RD and the idea of using LAFs for EG estimations can be applied at the same time. The 3LAF LO-RANSAC-RD algorithm estimates EG from three region-to-region correspondences as 3LAF LO-RANSAC, then uses the same LO step as LO-RANSAC-RD. The Courtyard QY problem Fig. 4 is solved by 3LAF LO-RANSAC-RD after less then 30 samples (compare with Tab. 1) with the same precision as when using LO-RANSAC-RD. A more complicated matching with only 28% of inliers (see Fig. 9) was solved after only 133 RANSAC cycles.

6. CONCLUSIONS

In this paper, the LO step to RANSAC algorithm was introduced. Three applications demonstrating the properties of LO-RANSAC framework were presented.

First, LO-RANSAC-RD, an algorithm for joint estimation of the epipolar geometry and radial distortion was presented. We showed, that the algorithm: 1. has the same complexity as the 7pt RANSAC, i.e. $\mathcal{O}(\varepsilon^{-7})$, 2. produces

Fig. 8: Epipolar lines and correspondences superimposed over 'Ascona' pair. Close-ups with regions and LAFs highlighted are shown in the middle.

more inliers than the 7pt LO-RANSAC and hence can be terminated earlier, and 3. is more stable than the 9pt RANSAC (both the number of detected inliers and the estimated parameter of radial distortion have smaller variance).

Second, 3LAF LO-RANSAC – a new algorithm for the correspondence problem – was described. Exploiting output of the processes proposed in [7] for computation of affine-invariant local frames, three point-to-point correspondences were found for each region-to-region correspondence and used in epipolar geometry estimation. We have experimentally shown that: 1. 3LAF LO-RANSAC estimates epipolar geometry in time that is orders of magnitude faster than the standard method, 2. that the precision of the 3LAF LO-RANSAC and the standard method are comparable, and 3. that RANSAC without the LO step applied to triplets of points from a single region is significantly less precise than the new 3LAF LO-RANSAC algorithm. The presented matching method is pushing the limit of solvable problems, allowing EG estimation in correspondence problems with the ratio of inliers below 10%.

Finally, the combination of the previous two algorithms was tested. The 3LAF LO-RANSAC-RD algorithm has advantages of both LO-RANSAC-RD and 3LAF LO-RANSAC. The simultaneous estimation of EG and RD increases the precision and the number of correct matches, the time complexity is reduced to $\mathcal{O}(\varepsilon^{-3})$. This compares favourably with the $\mathcal{O}(\varepsilon^{-9})$ complexity of the state-of-the-art.

7. REFERENCES

[1] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. In *Proc. DAGM*. Springer-Verlag, 2003.

[2] 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.

[3] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proc. of CVPR*, volume 1, pages 125–132, 2001.

[4] R. Hartley. In defence of the 8-point algorithm. In *ICCV95*, pages 1064–1070, 1995.

[5] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[6] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proc. of BMVC*, volume 1, pages 384–393. BMVA, 2002.

[7] J. Matas, Š. Obdržálek, and O. Chum. Local affine frames for wide-baseline stereo. In *Proc. ICPR*, volume 4, pages 363–366. IEEE CS, Aug 2002.

[8] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proc. ECCV*, volume 1, pages 128–142, 2002.

[9] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. ICCV*, pages 754–760, 1998.

[10] B. Tordoff and D. Murray. Guided sampling and consensus for motion estimation. In *Proc. 7th ECCV*, volume 1, pages 82–96. Springer-Verlag, 2002.

[11] P. Torr, A. Zisserman, and S. Maybank. Robust detection of degenerate configurations while estimating the fundamental matrix. *CVIU*, 71(3):312–333, Sep. 1998.

[12] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *CVIU*, 78:138–156, 2000.

[13] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proc. 11th BMVC*, 2000.