Randomized RANSAC *

Jiří Matas, Ondřej Chum

Center for Machine Perception, Faculty of Electrical Engineering

Czech Technical University, Technicka 2, Prague, Czech Republic

tel: +420 2 2435 7637, fax: +420 2 2435 7385

e-mail: [matas,chum]@cmp.felk.cvut.cz

Abstract

Many computer vision algorithms include a robust estimation step where model parameters are computed from a data set containing a significant proportion of outliers. The RANSAC algorithm is possibly the most widely used robust estimator in the field of computer vision. In the paper we show that under a broad range of conditions, RANSAC efficiency is significantly improved if *hypothesis evaluation is randomized*.

A new randomized (hypothesis evaluation) version of RANSAC, R-RANSAC, is introduced. Computational savings are achieved by typically evaluating only a fraction of data points for models contaminated with outliers. The idea is implemented in a two-step evaluation procedure. A mathematically tractable class of statistical preverification tests for test samples is introduced. For this class we derive an approximate relation for optimal setting of its single parameter. The proposed pre-test is verified on both synthetic data and real-world problems and a significant increase in speed is shown.

1 Introduction

Many computer vision algorithms include a robust estimation step where model parameters are computed from a data set containing a significant proportion of outliers. The RANSAC[1] algorithm introduced by Fishler and Bolles in 1981 [2] is possibly the most widely used robust estimator in the field of computer vision. RANSAC has been applied in the context of short baseline stereo [11, 13], wide baseline stereo matching [8, 14, 10], motion segmentation [11], mosaicing [6], detection of geometric primitives [1], robust eigenimage matching [4] and elsewhere.

The structure of the RANSAC algorithm is simple but powerful. Repeatedly, subsets are randomly selected from the input data and model parameters fitting the sample are

*The authors were supported by the European Union under project IST-2001-32184, the Czech Ministry of Education under project MSM 212300013 and by The Grant Agency of the Czech Republic under project GACR 102/02/1539.

[1] **RAN**dom **SA**mple **C**onsensus

computed. The size of the random samples is the smallest sufficient for determining model parameters. In a second step, 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 number of inliers, i.e. the number of data points consistent with the model. The process is terminated when the likelihood of finding a better model becomes low. The strength of the method stems from the fact that it is sufficient to select a single random sample not contaminated by outliers to find a good solution. Depending on the complexity of the model (the size of random samples) RANSAC can handle contamination levels well above 50%, which is commonly assumed to be a practical limit in robust statistics [9].

The speed of RANSAC depends on two factors. Firstly, the level of contamination determines the number of random samples that have to be taken to guarantee a certain confidence in the optimality of the solution. Secondly, the time spent evaluating the quality of each of the hypothesized model parameters is proportional to the size N of the data set. Typically, a very large number of erroneous model parameters obtained from contaminated samples are evaluated. Such models are consistent with only a small fraction of the data. This observation can be exploited to significantly increase the speed of the RANSAC algorithm.

As the main contribution of this paper, we show that under a broad range of conditions, RANSAC efficiency is significantly improved if *hypothesis evaluation is randomized*. The core idea of the Randomized (hypothesis evaluation) RANSAC is that most model parameter hypotheses evaluated are influenced by outliers. For such erroneous models, it is sufficient to test only a small number of data points d from the total of N points ($d \ll N$) to conclude, with high confidence, that they do not correspond to the sought solution. The idea is implemented in a two-step evaluation procedure. First, a statistical test is performed on d randomly selected data points. The final evaluation on all N data points is carried out only if the pre-test is passed. The increase in speed of the modified RANSAC depends on the likelihoods of the two types of errors made in the pre-test: 1. rejection of an uncontaminated model and 2. acceptance of a contaminated model. Since RANSAC is already a randomized algorithm, the randomization of model evaluation does not change the nature of the solution – it is only correct with a certain probability. However, the same confidence in the solution is obtained in, on average, a shorter time.

Finding an optimal pre-test with the fastest average behaviour is naturally desirable, but very complex. Instead we introduce in Section 3 a mathematically tractable class of pre-tests based on small test samples. For this class we derive an approximate relation for optimal setting of its single parameter. The proposed pre-tests are assessed on both synthetic data and real-world problems and performance improvements are demonstrated.

The structure of this paper is as follows. First, in Section 2, the concept of evaluation with pre-tests is introduced and formulae describing the total complexity of the algorithm are derived. Both the number of samples drawn and the amount of time spent on evaluation of a hypothesized model are discussed in detail. In Section 3, the *d-out-of-d* class of pre-test is introduced and analyzed. In Section 4 both simulated and real experiments are presented and their results discussed. The paper is concluded in Section 5 and plans for future work are discussed.

2 Randomized RANSAC

In this section, the time complexity of the RANSAC algorithm is expressed as a function of quantities that characterise the input data. We start the presentation by introducing the most important symbols used. The set of all data points is denoted U, the number of data points $N = |U|$, and ε represents the fraction of inliers contained in the data points. The symbol m is the size of the sample, i.e. the number of data points necessary to compute model parameters.

Let us first express the total time spent in the R-RANSAC procedure. The time needed to verify the consistency of one data point with the hypothesized parameters was chosen as a unit of time. The average time spent in R-RANSAC in number of verified data points is

$$J = k(\bar{t} + t_M), \tag{1}$$

where k is the number of samples drawn, $\bar{t}$ is the average number of verified data points within one model evaluation, and t_M is the time necessary to compute the parameter of the model from the selected sample. Note that t_M from (1) is a constant independent of both N and ε.

From (1) we can see, that the average time spent in R-RANSAC depends on both the number of samples drawn k and the average time required to process each sample. The analysis of these two components follows.

The number of tested hypothesis, which is equal to the number of samples, depends on the termination condition. Two different termination criteria may be adopted in RANSAC. The hypothesize-verify loop is either stopped after evaluation of more samples than expected on average before a good (uncontaminated) sample is selected or, alternatively, the number of samples is chosen to ensure that the probability that a better-than-currently-best sample is missed is lower than a predefined confidence level. We show that the stopping times for the two cases, average-driven and confidence-driven, differ only by a multiplicative factor and hence the optimal value in the proposed test is reached with the same parameters.

Since the sample is selected without repetition, the probability of taking a good sample is

$$P_I = \frac{\begin{pmatrix} I \\ m \end{pmatrix}}{\begin{pmatrix} N \\ m \end{pmatrix}} = \frac{I!\,(N-m)!}{(I-m)!\,N!} = \prod_{j=0}^{m-1} \frac{I-j}{N-j},$$

where $I = \varepsilon N$ stands for the number of inliers. For $N \gg m$ a simple and accurate approximation is obtained

$$P_I \approx \varepsilon^m, \tag{2}$$

which is exactly correct for sampling with repetition. The average number of samples taken before the first uncontaminated is given by (from properties of the geometric distribution)

$$\bar{k} = \frac{1}{\varepsilon^m \alpha}. \tag{3}$$

Note that for the randomised version of RANSAC the number of samples is higher than or equal to the standard version, because a valid solution may be rejected in a preliminary test

In:	$U = \{x_i\}$	set of data points, $\lvert U\rvert = N$
	$f : S \to p$	computes model parameters from a data point sample
	$\rho(p, x)$	the cost function for a single data point
Out:	p^*	parameters of the model maximizing the cost function

$k := 0$

Repeat until P{better solution exists} $< \eta$
(a function of C^* and no. of steps k)

$k := k + 1$

I. Hypothesis

(1) select randomly set $S_k \subset U$, $|S_k| = m$
(2) compute parameters $p_k = f(S_k)$

II. Preliminary test

(3) perform test based on $\mathrm{d} \ll N$ *data points*
(4) continue verification only if the test is passed

III. Evaluation

(5) compute cost $C_k = \sum_{x \in U} \rho(p_k, x)$
(6) if $C^* < C_k$ then $C^* := C_k$, $p^* := p_k$

Table 1: *Summary of* RANSAC *and* R-RANSAC *algorithms. The step II is added to* RANSAC *to randomize its evaluation.*

with probability $1 - \alpha$. In the confidence-driven sampling, at least k samples have to be taken to reduce the probability of missing a good sample below a predefined confidence level η. Thus we get, as in [11]),

$$\eta = (1 - \varepsilon^m \alpha)^k, \tag{4}$$

and solving for k leads to

$$k = \frac{\log \eta}{\log (1 - \varepsilon^m \alpha)}. \tag{5}$$

Since $(1 - x)$ is the first order Taylor expansion of e^{-x} at zero, and $(1 - x) \leq e^{-x}$, we have

$$\begin{aligned} \eta \;=\; (1 - \varepsilon^m \alpha)^k &\leq e^{-\varepsilon^m \alpha k} \\ \ln \eta &\leq -\varepsilon^m \alpha k \\ \frac{-\ln \eta}{\varepsilon^m \alpha} &\geq k \end{aligned}$$

We see, that $k \leq \bar{k}(-\ln \eta)$, where $-\ln \eta$ is a predefined constant, so all formulae obtained for the η-confidence driven case can be trivially modified to cover the average case.

The Number of data points points tested. So far we have seen that introduction of a preliminary test has *increased the number of samples drawn*. For the pre-test to make sense, this effect must be more than offset by the reduction in the average number of data points tested per hypothesis. There are two cases to be considered. First, with probability P_I an uncontaminated ('good') sample is drawn. Then the preverification test is passed with probability α and all N data points are verified. Else, with probability $1-\alpha$, this good sample is rejected and only $\bar{t}_\alpha$ data points are on average tested. In the second case, a contaminated ('bad') sample is drawn, and this happens with probability $1-P_I$. Again either the pre-verification step is passed, but this time with a *different probability* β, and the full test with all N data points is carried out, or with probability $1-\beta$, only $\bar{t}_\beta$ data points are tested in the preverification test.

Here β stands for the probability, that a bad sample passes the preverification test. Note that it is important that $\beta \ll \alpha$, i.e. a bad (contaminated) sample is consistent with a smaller number of data points than a good sample. Forming a weighted average of the four cases, the formula for the average number of tests per sample is obtained:

$$\bar{t}(d) = P_I(\alpha N + (1-\alpha)\bar{t}_\alpha) + (1-P_I)(\beta N + (1-\beta)\bar{t}_\beta). \tag{6}$$

Values of α, β, $\bar{t}_\alpha$, and $\bar{t}_\beta$ depend on the type of preverification test.

3 The $T_{d,d}$ Test

In this section we introduce a simple and thus mathematically tractable class of preverification tests. Despite its simplicity, we show in the simulations and experiments of Section 4 its potential. The test we will analyze is defined as follows: Pass the test if all d data points out of d randomly selected are consistent with the hypothesized model. In the rest of this section we derive the optimal value for d. But first of all we have to derive constants introduced in the previous section

$$\alpha = \varepsilon^d \qquad \text{and} \qquad \beta = \delta^d,$$

where δ is the probability that a data point is consistent with a "random" model. Since we do not need to test all d, just to find first failure, the average time spent in the preverification test is

$$\bar{t}_\alpha = \sum_{i=1}^{d} i\,(1-\varepsilon)\,\varepsilon^{i-1} \qquad \text{and} \qquad \bar{t}_\beta = \sum_{i=1}^{d} i\,(1-\delta)\,\delta^{i-1}$$

Since

$$\sum_{i=1}^{d} i(1-x)x^{i-1} \le \sum_{i=1}^{\infty} i(1-x)x^{i-1} = \frac{1}{1-x}, \tag{7}$$

we have

$$\bar{t}_\alpha \le \frac{1}{1-\varepsilon} \qquad \text{and} \qquad \bar{t}_\beta \le \frac{1}{1-\delta}.$$

The relationship we get after substituting these into (6)

$$\bar{t}(d) \approx \varepsilon^m \left(\varepsilon^d N + \frac{1-\varepsilon^d}{1-\varepsilon}\right) + (1-\varepsilon^m)\left(\delta^d N + \frac{1-\delta^d}{1-\delta}\right)$$

is too complicated, so we incorporate the following approximations

$$(1-\varepsilon^m)\frac{1-\delta^d}{1-\delta} \approx 1,$$

$$(1-\varepsilon^m)\delta^d N \approx \delta^d N, \text{ and}$$

$$\varepsilon^d N \gg \frac{1-\varepsilon^d}{1-\varepsilon}.$$

After applying these approximations, we have

$$\bar{t}(d) \approx N\,\delta^d + 1 + \varepsilon^{m+d}\,N \tag{8}$$

The average time spent in R-RANSAC in number of verified data points is then approximately

$$J(T_{d,d}) \approx \frac{1}{\varepsilon^m\,\varepsilon^d}\left(N\,\delta^d + \varepsilon^{m+d}\,N + 1 + t_M\right) \tag{9}$$

We are looking for the minimum of $J(T_{d,d})$ and hence we can solve $\frac{\partial J(T_{d,d})}{\partial d} = 0$ for d and we get optimal length of the $T_{d,d}$ test as follows

$$d^* \approx \frac{\ln\left(\frac{\ln\varepsilon(t_M+1)}{N\,(\ln\delta - \ln\varepsilon)}\right)}{\ln\delta}. \tag{10}$$

The value of d_{opt} must be an integer greater or equal to zero, so it could be obtained as

$$d_{opt} = \max(0, \arg\min_{d\in\{\lfloor d^*\rfloor, \lceil d^*\rceil\}} J(T_{d,d})). \tag{11}$$

Since the cost function $J(T_{d,d})$ has only one extreme and for $d \rightarrow \pm\infty$ we have $J(T_{d,d}) \rightarrow \infty$, we can say that R-RANSAC is faster than the standard RANSAC if

$$J(T_{0,0}) > J(T_{1,1}).$$

From this equation we get

$$N > (t_M + 1)\frac{1-\varepsilon}{\varepsilon - \delta}. \tag{12}$$

4 Experiments

In this section are experiments that show the usefulness of the new randomised RANSAC algorithm with a preverification tests. The speed-up is demonstrated on the problem of epipolar geometry estimation. Three experiments are conducted on data from a synthetic, a short (standard) and wide-baseline stereo matching problem. Results of these experiments are summarized in tables 2, 3, and 4 respectively. The structure of the tables is the following. The first column shows the length d of the $T_{d,d}$ test, where $d = 0$ means standard RANSAC. The number of samples, each consisting of 7 point-to-point correspondences, that were used for model parameter estimation is given in the second column. Since the seven-point algorithm [3] for computation of the fundamental matrix may lead to one or three solutions, the next column, labeled 'models', shows the number of hypothesized

d	samples	models	tests	inliers	time
0	1866	4569	6821218	600	25.0
1	4717	11536	16311	600	6.0
2	11849	28962	33841	600	15.1

Table 2: *Synthetic experiment on 1500 correspondences, 40% of inliers, 30 repetitions.*

d	samples	models	tests	inliers	time
0	480	1146	766875	343	2.6
1	960	2301	83953	342	1.4

Table 3: *Short baseline experiment on 676 tentative correspondences.*

fundamental matrices. The 'tests' column displays the number of point-to-point correspondences evaluated during the procedure. In the penultimate column, the average number of inliers detected is given. The last column is rather informative and shows the time in seconds taken by the algorithm. This is strongly dependent on the implementation.

Synthetic experiment. 1500 correspondences were generated, 900 outliers and 600 inliers. Since the run-time of both RANSAC and R-RANSAC is a random variable, the programs were executed 30 times and averages were taken. Result are shown in the Table 2. Since the number of correspondences is large, the standard RANSAC algorithm spends a long time verifying all correspondence as can be seen in column 'tests'.

Short baseline experiment was conducted on the images from a standard dataset of the Leuven castle [7]. There were 676 tentative correspondences formed by the Harris interest points followed by cross-correlation of its neigbourhood. The tentative correspondences contained approximately 60% of inliers. Looking at Table 3 we see that approximately twice as many fundamental matrices were hypothesized in R-RANSAC, but more than nine times less correspondences were evaluated.

Wide baseline experiment on the BOOKSHELF dataset. The tentative correspondences were formed as follows. Discriminative regions (MSERs, SECs) [5] were detected. Robust similarity functions on the affine invariant description were used to establish the mutually nearest pair of regions. Point correspondences were obtained as centers of gravity of those regions. There were less then 40% of inliers among the correspondences.

5 Conclusion

In this paper, we presented a new R-RANSAC algorithm, which increased the speed of model parameter estimation under a broad range of conditions, due to the hypothesis evaluation

d	samples	models	tests	inliers	time
0	3094	7582	3078184	161	12.9
1	6366	15583	178217	164	8.7

Table 4: *Wide baseline experiment on 413 tentative correspondences.*

Figure 1: *Short baseline image set*

Figure 2: *Wide baseline image set*

being randomized. For samples contaminated by outliers, it was shown that it was sufficient to test only a small number of data points $d \ll N$ to conclude with high confidence that they do not correspond to the sought solution. The idea was implemented in a two-step evaluation procedure (Table 1). We also introduced in Section 3 a mathematically tractable class of pre-tests based on small test samples. For this class we derived an approximate relation for optimal setting of its single parameter. The proposed pre-test was verified on both synthetic data and real-world problems and a significant increase in speed was observed. The task for the future is to design the optimal preverification test.

References

[1] J. Clarke, S. Carlsson, and A. Zisserman. Detecting and tracking linear features efficiently. In *Proc. 7th BMVC*, pages 415–424, 1996.

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

[4] A. Leonardis, H. Bischof, and R. Ebensberger. Robust recognition using eigenimages. Technical Report TR-47, PRIP, TU Wien, 1997.

[5] Jiří Matas, Ondřej Chum, Martin Urban, and Tomáš Pajdla. Wide-baseline stereo from distinguished regions. Research Report CTU–CMP–2001–33, Center for Machine Perception, K333 FEE Czech Technical University, Prague, Czech Republic, November 2001.

[6] P. McLauchlan and A. Jaenicke. Image mosaicing using sequential bundle adjustment. In *Proc. BMVC*, pages 616– 62, 2000.

[7] M. Pollefeys. *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. PhD thesis, ESAT-PSI, K.U.Leuven, 1999.

[8] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proc. International Conference on Computer Vision*, pages 754–760, 1998.

[9] Peter J. Rousseeuw and Annick M. Leroy. *Robust Regression and Outlier Detection*. Wiley, 1987.

[10] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *Proc. 8th International Conference on Computer Vision, Vancouver, Canada*, July 2001.

[11] P. H. S. Torr. *Outlier Detection and Motion Segmentation*. PhD thesis, Dept. of Engineering Science, University of Oxford, 1995.

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

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

[14] Tinne Tuytelaars and Luc Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proc. 11th British Machine Vision Conference*, 2000.