

Uncertain RanSaC

Ben Tordoff and Roberto Cipolla
 Department of Engineering, University of Cambridge
 Trumpington Street, Cambridge CB2 1PZ, UK

Abstract

This paper describes a new enhancement to the technique of estimation by Random Sampling and Consensus. The current state-of-the-art random sampling schemes are discussed, in particular looking at the speed of discovery of the solution. An improved support function is examined, using linear approximations to the second moments of the parameter PDFs to more accurately propagate noise. This improves the proportion of hypotheses which find the underlying motion.

1 Introduction

Robust estimation of a signal from highly contaminated data was revolutionised by the idea of estimation using Random Sampling And Consensus (RanSaC) introduced by Fischler and Bolles [4] and later reaffirmed in the statistical literature by Rousseeuw and Leroy [10]. Although its utility is widespread (two appropriately random examples of substantially different uses are [7], [9]), RanSaC has proved particularly effective for estimating inter-image transforms such as the homography and fundamental matrix [14, 11, 16, 15, 13], where the presence of mis-matches in the input set of image feature correspondences can, and usually does, cause havoc.

In the original scheme a minimal set of data points are selected randomly, used to linearly calculate the transform and support then sought in the remaining data using a binary decision boundary on the transformation error. Torr and Zisserman [13] improved on this by replacing the discrete support threshold by a continuous cost function derived from a Gaussian plus uniform assumption about the inlying and clutter processes. The support score produced is then related to the likelihood of the hypothesis, resulting in a maximum likelihood estimator, MLESaC.

Tordoff and Murray [12] further refined this algorithm by showing that priors on the data points both increase efficiency and produce posterior estimates of which data points are valid and which are clutter. In addition they replaced the random sampling with a sampling guided by the priors, biasing the search towards valid hypotheses.

Further reductions in computational cost are also gained using the complementary pre-emptive approach of Nister

[8] and others. This allows poor motion hypotheses to be discarded early, reducing the cost of hypothesis evaluation.

Whilst MLESaC improved accuracy and Guided Sampling and Consensus (GuiSaC) and pre-emption increased efficiency, the problem of knowing when sampling could be stopped remained open. It is well known (eg. [1, 2, 12]) that the usual calculation based on the proportion of valid data was often dramatically over-optimistic. Chum *et al* [1, 2] addressed this problem by resampling from the set of inlying data when a well supported hypothesis is found. This explores the local parameter space around good solutions and widens the reach of “nearly correct” hypotheses.

Here we also examine the inadequacies of the stopping criterion, but do so by modelling the uncertainties in the estimation process explicitly using Gaussian assumptions and linearisations. In particular section 3 looks at how existing results on uncertain multiview geometry can be used to better model estimation errors. Section 4 compares the performance of this new scheme with the standard model before conclusions are drawn. First, however, the basic guided sampling and consensus scheme is described in detail.

2 Review: Guided Sampling and Consensus

GuiSaC [12] pre-supposes that feature detection and matching processes have been applied to successive frames in an image sequence, yielding an input data set \mathcal{C} of n feature correspondences.

Each feature typically has several putative matches, the winner chosen based on a match “score”. In the case of correlation-based matching this score is the likelihood of each correspondence being in error, determined from learnt histograms of normalised correlation scores for correct and incorrect matching [12]. Similarly, for matching based on more esoteric descriptors (such as Lowe’s SIFT operator [6]) histograms against any appropriate similarity measure can be used (eg. angle between description vectors) again leading to the probability of validity given score $p(v_i|s_i)$ (abbreviated to $p(v_i)$ below).

These probabilities are used as priors to the motion estimation process, weighting the random selection of correspondences. For brevity we will consider only two view relations, in particular the planar homography H (4 correspondences per sample) and the fundamental matrix F (using an

8 correspondence method).

Each motion hypothesis M_h is scored according to how well it predicts the observed correspondences. If a correspondence is clutter (with probability $1 - p(v_i)$) its error under the hypothesised motion, r_{hi} , is assumed uniform. If, however, it supports the motion it should be correctly mapped up to Gaussian noise in the image measurements. This gives a likelihood which is a Gaussian plus Uniform mixture:

$$p(r_{hi}|M_h) = \left[\frac{e^{-r_{hi}^2/2\sigma^2}}{(2\pi\sigma^2)^{d/2}} \right] p(v_i) + \left[\frac{1}{A_i} \right] (1 - p(v_i)), \quad (1)$$

where σ is the measurement uncertainty (image noise), A_i is the search area (often the same for all features) and d the dimensionality of the error — 1D for the fundamental matrix and 2D for homographies. Making an assumption of independence, the overall probability of the correspondence set given the motion hypothesis is

$$p(C|M_h) = p(R_h|M_h) = \prod_{i=1}^n p(r_{hi}|M_h),$$

and maximising this over many samples eventually yields a maximum likelihood estimate of the motion. Posterior estimates of each correspondence's validity, $p(v_i|M_h, s_i)$, are also produced and can be used to weight a subsequent linear or nonlinear estimation.

2.1 Stopping criterion

Summing the posterior validity likelihoods of the correspondences gives an estimate of the expected proportion of valid matches $E(\alpha) \approx \frac{1}{n} \sum_i^n p(v_i|M_h, s_i)$ and the proportion from the best motion hypothesis is used to estimate the confidence that at least one sample has contained no clutter

$$p(M_c) = 1 - (1 - E(\alpha)^m)^I, \quad (2)$$

where m is the sample size and I the number of samples taken so far. Sampling can therefore be stopped once the confidence has risen to a suitable level (eg. 99%).

An example pair of images and the inliers ($p(v_i|M_h, s_i) > 50\%$) and outliers when GuiSaC'ing for a planar homography are shown in figure 1. For these images the total number of valid correspondences has been manually determined. Running GuiSaC for many thousands of iterations, figure 2 shows the proportion of the valid set discovered against the proportion of trials able to find them. In this case the expected proportion of valid hypotheses is approximately $\alpha^m = 0.5\%$, but the RanSaC curve suggests that only 0.33% find half or more of the inliers. This is alleviated by guiding the sampling so that more valid hypotheses are tried, but many valid hypotheses still fail to find many of the inliers. This is for two reasons:

1. **ill-conditioning** — even valid correspondences sometimes badly constrain the motion

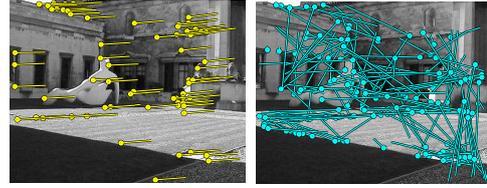


Figure 1. GuiSaC at work on an image pair related by a homography. The correspondences are separated into inliers (left) and outliers (right) using their posterior probability of belonging to the best motion hypothesis.

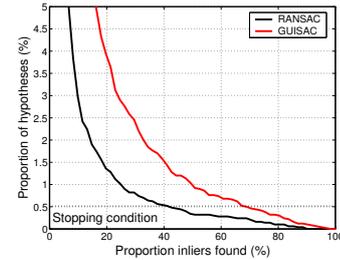


Figure 2. The proportion of inliers found by each motion hypothesis gives an indication of the performance of the algorithms. The stopping condition predicts that around 1 in 200 hypotheses will be valid, but in these trials only 1 in 300 finds half or more of the inlier set.

2. **noise** — the motion is hypothesised from noisy measurements and is therefore noisy.

These obstacles can both be better understood by estimating the uncertainty in the motion parameters.

3 Uncertain motion estimation

The two view relations considered here (homography and fundamental matrix) can both be estimated as the null vector of a symmetric measurement matrix. For a motion with parameter vector \mathbf{t} we build the system

$$\mathbf{t}^\top \mathbf{A}^\top \mathbf{A} \mathbf{t} = 0$$

such that $\mathbf{t} \approx \mathbf{u}_0$, the eigenvector of $\mathbf{M} = \mathbf{A}^\top \mathbf{A}$ corresponding to the smallest eigenvalue λ_0 ($= 0$ when a minimal set of data is used). The remaining eigenvectors describe the relationship between perturbations of the measurement matrix and the estimated transform (see eg. [5]),

$$\delta \mathbf{u}_0 \approx \sum_{i=1}^n \frac{\mathbf{u}_i \mathbf{u}_i^\top}{\lambda_0 - \lambda_i} \delta \mathbf{M} \mathbf{u}_0 = \mathbf{J} \delta \mathbf{M} \mathbf{u}_0,$$

The covariance of \mathbf{u}_0 , and hence of the transform parameters is calculated as

$$\begin{aligned} \mathbf{C}_T &= \mathbf{E} [\delta \mathbf{u}_0 \delta \mathbf{u}_0^\top] = \mathbf{E} [\mathbf{J} \delta \mathbf{M} \mathbf{u}_0 \mathbf{u}_0^\top \delta \mathbf{M}^\top \mathbf{J}^\top] \\ &= \mathbf{J} \mathbf{E} [\mathbf{A}^\top \delta \mathbf{A} \mathbf{u}_0 \mathbf{u}_0^\top \delta \mathbf{A}^\top \mathbf{A}] \mathbf{J}^\top. \end{aligned}$$

where $\delta\mathbf{M} \approx \delta\mathbf{A}^\top \mathbf{A} + \mathbf{A}^\top \delta\mathbf{A}$. If the noise on each feature is assumed independent, and each correspondence gives m rows of \mathbf{A} , then some rearrangement gives the contribution from each feature to be

$$\mathbf{C}_{T_i} = \mathbf{J}\mathbf{C}_{A_i}\mathbf{J}^\top \quad \text{with} \quad \mathbf{C}_{A_i} = \sum_{j=1}^m \sum_{k=1}^m (\mathbf{a}_j \mathbf{a}_k^\top) \mathbf{u}_0^\top \mathbf{E}[\delta\mathbf{a}_j \delta\mathbf{a}_k^\top] \mathbf{u}_0,$$

and the total covariance of the transform is estimated from N correspondences as $\mathbf{C}_T = \sum_{i=1}^N \mathbf{C}_{T_i}$. The uncertainty in the error r_i now arises from three sources:

1. Measurement uncertainty in image 1, \mathbf{C}_X
2. Uncertainty in the estimated transform, \mathbf{C}_T
3. Measurement uncertainty in image 2, $\mathbf{C}_{X'}$

Assuming image noise is isotropic and Gaussian, the homogeneous covariances are $\mathbf{C}_X = \text{diag}(\sigma^2, \sigma^2, 0)$ and $\mathbf{C}_{X'} = \text{diag}(\sigma'^2, \sigma'^2, 0)$, and the expressions for \mathbf{C}_{A_i} can be simplified as follows.

3.1 Fundamental matrix

Each correspondence produces a single row for the measurement matrix \mathbf{A}

$$\mathbf{a}^\top = (\mathbf{x}^\top x' \quad \mathbf{x}^\top y' \quad \mathbf{x}^\top) \quad \text{where} \quad \mathbf{x} = (x, y, 1)^\top.$$

Assuming isotropic Gaussian noise of variance σ^2 in image 1 and σ'^2 in image 2,

$$\mathbf{C}_{A_i} = \gamma \mathbf{a} \mathbf{a}^\top$$

$$\text{where} \quad \gamma = \sigma^2 [(x'f_1 + y'f_4 + f_7)^2 + (x'f_2 + y'f_5 + f_8)^2] + \sigma'^2 [(xf_1 + yf_2 + f_3)^2 + (xf_4 + yf_5 + f_6)^2].$$

The total covariance of the epipolar line \mathbf{l} is then

$$\mathbf{C}_l = \mathbf{F}\mathbf{C}_X\mathbf{F}^\top + \mathbf{B}\mathbf{C}_T\mathbf{B}^\top \quad \text{where} \quad \mathbf{B} = \begin{bmatrix} \mathbf{x}^\top & \mathbf{0}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{x}^\top & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{x}^\top \end{bmatrix},$$

and both the line and its covariance are then normalised such that the distance $d = \mathbf{l}^\top \mathbf{x}'$. Adding the covariance of the target point $\mathbf{C}_{X'}$, the distance of the point from the line has uncertainty

$$\sigma_d^2 = \mathbf{l}^\top \mathbf{C}_{X'} \mathbf{l} + \mathbf{x}'^\top \mathbf{C}_l \mathbf{x}'.$$

3.2 Homography

For a homography [3], each feature correspondence provides three rows¹ to the measurement matrix

$$\begin{pmatrix} \mathbf{a}_x^\top \\ \mathbf{a}_y^\top \\ \mathbf{a}_z^\top \end{pmatrix} = \begin{pmatrix} \mathbf{0}_3 & -\mathbf{x}^\top & y' \mathbf{x}^\top \\ \mathbf{x}^\top & \mathbf{0}_3 & -x' \mathbf{x}^\top \\ -y' \mathbf{x}^\top & x' \mathbf{x}^\top & \mathbf{0}_3 \end{pmatrix},$$

¹Note that $\mathbf{a}_z = -x' \mathbf{a}_x - y' \mathbf{a}_y$ and the three rows only provide two constraints. For all normal geometries the third row and all terms involving it may be safely ignored, but are included here for completeness.

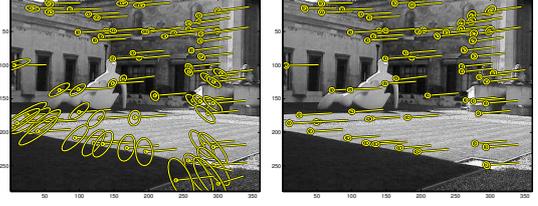


Figure 3. Examples of the covariances in image 2, for inliers to two motion hypotheses. Even for good hypotheses such as these using the full uncertainty finds more inliers than would otherwise be detected.

with $\mathbf{x} = (x, y, 1)^\top$. Again, assuming independent Gaussian noise:

$$\mathbf{C}_{A_i} = \gamma_{xx}(\mathbf{a}_x \mathbf{a}_x^\top) + \gamma_{yy}(\mathbf{a}_y \mathbf{a}_y^\top) + \gamma_{zz}(\mathbf{a}_z \mathbf{a}_z^\top) - \gamma_{xy}(\mathbf{a}_x \mathbf{a}_y^\top + \mathbf{a}_y \mathbf{a}_x^\top) - \gamma_{xz}(\mathbf{a}_x \mathbf{a}_z^\top + \mathbf{a}_z \mathbf{a}_x^\top) - \gamma_{yz}(\mathbf{a}_y \mathbf{a}_z^\top + \mathbf{a}_z \mathbf{a}_y^\top)$$

where

$$\begin{aligned} \gamma_{xx} &= \sigma^2 \{ (h_4 - h_7 y')^2 + (h_5 - h_8 y')^2 \} + \sigma'^2 \{ (h_7 x + h_8 y + h_9)^2 \} \\ \gamma_{xy} &= \sigma^2 \{ (h_1 - h_7 x')(h_4 - h_7 y') + (h_2 - h_8 x')(h_5 - h_8 y') \} \\ \gamma_{yy} &= \sigma^2 \{ (h_2 - h_8 x')^2 + (h_1 - h_7 x')^2 \} + \sigma'^2 \{ (h_7 x + h_8 y + h_9)^2 \} \\ \gamma_{xz} &= \sigma^2 \{ (h_4 x' - h_1 y')(h_4 - h_7 y') + (h_5 x' - h_2 y')(h_5 - h_8 y') \} \\ &\quad + \sigma'^2 \{ (h_1 x + h_2 y + h_3)(h_7 x + h_8 y + h_9) \} \\ \gamma_{yz} &= \sigma^2 \{ (h_4 x' - h_1 y')(h_7 x' - h_1) + (h_5 x' - h_2 y')(h_8 x' - h_2) \} \\ &\quad + \sigma'^2 \{ (h_4 x + h_5 y + h_6)(h_7 x + h_8 y + h_9) \} \\ \gamma_{zz} &= \sigma^2 \{ (h_1 y' - h_4 x')^2 + (h_2 y' - h_5 x')^2 \} \\ &\quad + \sigma'^2 \{ (h_1 x + h_2 y + h_3)^2 + (h_4 x + h_5 y + h_6)^2 \}. \end{aligned}$$

The homogeneous covariance of the transferred point is

$$\mathbf{C} \approx \mathbf{H}\mathbf{C}_{x1}\mathbf{H}^\top + \mathbf{B}\mathbf{C}_T\mathbf{B}^\top$$

with \mathbf{B} as above. This is then linearised around the transferred point $(\hat{x}, \hat{y}, \hat{w})^\top = \mathbf{H}\mathbf{x}$ and the covariance in the second image added

$$\check{\mathbf{C}} = \mathbf{P}\mathbf{C}\mathbf{P}^\top + \begin{bmatrix} \sigma'^2 & 0 \\ 0 & \sigma'^2 \end{bmatrix} \quad \text{where} \quad \mathbf{P} = \frac{1}{\hat{w}^2} \begin{bmatrix} \hat{w} & 0 & -\hat{x} \\ 0 & \hat{w} & -\hat{y} \end{bmatrix}.$$

4 Uncertain RanSaC

We can now use these more detailed covariances in the Gaussian which represents the inlier likelihood in eqn. 1. Example covariances are shown in figure 3.

To analyse the benefits of this approach, we consider not just the number of inliers found, but whether these inliers are sufficient to accurately determine the motion. To evaluate this, each hypothesis is optimised using all discovered inliers, and the RMS error in the optimised motion parameters (elements of \mathbf{H} or \mathbf{F}) measured.

For the image pair of figure 1, histograms of the motion error for valid hypotheses (ie. those containing only valid

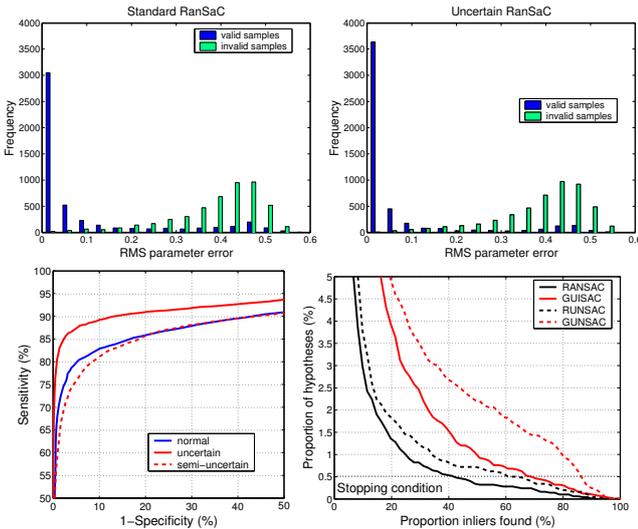


Figure 4. Histograms of the RMS motion error when estimated from inliers to traditional likelihood estimation (top-left) and uncertain estimation (top-right). The improvement is more clearly seen from an ROC curve (bottom-left) as the threshold error is varied (also shown is the improvement if just the point covariances are considered). This corresponds to an increased proportion of inliers (bottom-right).

correspondences) and invalid hypotheses are shown in figure 4, both for the standard method and for the uncertain motion estimation (whether or not the sampling is guided is irrelevant). More of the valid hypotheses converge to low motion error when the full covariances are used, most clearly seen from the ROC curve. This is due to the increased number of inliers that this method finds given a valid hypothesis — seen from the proportion plot.

As a modification of standard MLESaC (denoted RUNSaC), modelling the uncertainty makes the stopping criteria more realistic, and when combined with guided sampling (denoted GUnSaC) this stopping criteria can be exceeded. In both cases the accuracy of the resulting weighted linear solution is improved. This improvement comes at a significant computational cost — generating each hypothesis (and its covariance) is around 75% more costly, and since each feature has a different error covariance the cost of evaluation rises by about 150%. Using a pre-emptive scheme to avoid unnecessary evaluations is therefore important.

5 Conclusions

In this paper the linear propagation of noise covariances for motion estimation using RanSaC has been studied. Better handling of these error sources leads to the majority of valid motion hypotheses finding most of the valid correspondences. In turn this leads to fewer motion hypotheses being required for a given confidence level, or greater confidence for the same number of trials. This approach is com-

plementary to both the Guided Sampling and Pre-emptive methods, and when used in combination is an ideal basis for robust frame-rate tracking and structure from motion.

Acknowledgements

This work is part of the Cambridge MIT Institute “Visual Tracking for Modelling and Interaction” project.

References

- [1] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. In *DAGM Symposium on Pattern Recognition*, pages 236–243, 2003.
- [2] O. Chum, J. Matas, and S. Obdrzalek. Enhancing ransac by generalized model optimization. In *Proceedings of the 6th Asian Conference on Computer Vision*, 2004.
- [3] A. Criminisi, I. Reid, and A. Zisserman. A plane measuring device. *Image and Vision Computing*, 17(8):625–634, 1999.
- [4] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981.
- [5] G.H. Golub and C.F. Van Loan, editors. *Matrix Computations*. North Oxford Academic, 1983.
- [6] D. Lowe. Distinctive image features from scale invariant keypoints. *Int. Journal of Computer Vision*, 2(60):91–110, 2004.
- [7] B. Micusik and T. Pajdla. Using RANSAC for omnidirectional camera model fitting. In *Proc 8th Computer Vision Winter Workshop, Valtice, Czech Republic, February*, 2003.
- [8] D. Nister. Pre-emptive ransac for live structure and motion estimation. In *Proc 9th Int Conf on Computer Vision, Nice*, pages 199–206, 1996.
- [9] T. Okabe and Y. Sato. Object recognition based on photometric alignment using RANSAC. In *Proc of the IEEE Conf on Computer Vision and Pattern Recognition*, volume 1, pages 221–228, 2003.
- [10] P.J. Rousseeuw and A.M. Leroy. *Robust regression and outlier detection*. Wiley, New York, 1987.
- [11] L. Shapiro. *Affine analysis of image sequences*. Cambridge University Press, Cambridge, UK, 1995.
- [12] B. Tordoff and D.W. Murray. Guided sampling and consensus for motion estimation. In *Proc 7th European Conf on Computer Vision, Copenhagen*, pages 82–98, June 2002.
- [13] 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.
- [14] P.H.S. Torr and D.W. Murray. Statistical detection of independent movement from a moving camera. *Image and Vision Computing*, 11(4):180–187, 1993.
- [15] P.H.S. Torr and D.W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *Int. Journal of Computer Vision*, 24(3):271–300, September 1997.
- [16] G. Xu and Z. Zhang. *Epipolar geometry in stereo, motion and object recognition*. Kluwer Academic Publishers, 1996.