

# An Effective Bail-out Test for RANSAC Consensus Scoring

David Capel  
2d3 Ltd.  
14 Minns Business Park, Oxford OX2 0JB  
first.last@2d3.com

## Abstract

The *random sample consensus* (RANSAC) algorithm, along with its many cousins such as MSAC and MLESAC, has become a standard choice for robust estimation in many computer vision problems. Recently, a raft of modifications to the basic RANSAC algorithm have been proposed aimed at improving its efficiency. Many of these optimizations work by reducing the number of hypotheses that need to be evaluated. This paper proposes a complementary strategy that aims to reduce the average amount of time spent computing the consensus score for each hypothesis. A simple statistical test is proposed that permits the scoring process be terminated early, potentially yielding large computational savings. The proposed test is simple to implement, imposes negligible computational overhead, and is effective for any given size of data set. The approach is evaluated by estimation of the fundamental matrix for a large number of image pairs and is shown to offer a significant reduction in computational cost compared to recently proposed RANSAC modifications.

## 1 Introduction

The RANdom SAMple Consensus algorithm, proposed by Fischler and Bolles [5], is a simple and powerful method for estimating model parameters given a data set heavily contaminated by outliers to the correct model. It has proved extremely effective in sifting multi-view relations, such as homographies, epipolar geometry and the trifocal tensor, from point-to-point correspondences [10, 11, 6].

Recently, a number of modifications to the basic RANSAC algorithm have been proposed aimed at improving its overall efficiency. Tordoff and Murray [9] propose a scheme in which the similarity score associated with each putative feature correspondence is used to bias the sampling procedure to favour high-confidence matches and to eschew ambiguous matches, the motivation being that this will tend to sample good basis sets earlier in the process. Chum *et al.* [2, 1] proposed the LO-RANSAC variant in which a model optimization step is performed using the complete set of inliers whenever a new best hypothesis is discovered. This tends to improve the consensus score more rapidly than is the case in “vanilla” RANSAC, and hence the condition for termination may be reached more quickly. Chum *et al.* [3, 4] and Werner and Pajdla [13] propose a cheirality test for the fundamental matrix based on consideration of the *oriented* projective geometry

that allows hypotheses that do not satisfy the oriented epipolar constraint to be rejected without further evaluation. Nister [8] proposes a radically different approach in which multiple hypotheses are scored *in parallel*, with the least promising hypotheses being dropped at successive stages. This is intended for use in real-time tracking applications, in which a reasonable RANSAC hypothesis must be found within a *fixed* amount of time, but requires that the approximate inlier percentage is known *a priori*.

The algorithm proposed in this paper is not intended for real-time use, but rather it attempts to minimize the average time spent evaluating hypotheses in an off-line RANSAC process. No prior knowledge of the inlier percentage is required. We propose a novel test that may be used to “bail-out” part-way through the consensus scoring loop for a given hypothesis. Bail-out occurs when it becomes sufficiently certain that the score being computed will fail to better the previously recorded best consensus score. The test is simple to implement and imposes negligible computational overhead. This is most similar to the approach proposed by Matas and Chum [7] who propose a *d-out-of-d* ( $T_{d,d}$ ) pre-evaluation test that aims to reduce the average time spent computing consensus scores by insisting that all correspondences in a randomly selected subset of size  $d$  must be inliers to the given hypothesis in order for full consensus scoring to take place.

## 2 Background

The steps in the basic RANSAC algorithm are generally well known and understood, but for the sake of clarity they are detailed in table 1. It is assumed that we have a data set consisting of putative feature correspondences, a subset of which are consistent with (inliers to) some multi-view relation, such as a homography or epipolar relation, the parameters of which are unknown. The task is to estimate these parameters, along with the set of consistent correspondences (inliers).

In our experiments the data are point-to-point feature matches between two views and the relation considered is the *fundamental matrix*  $F$ . RANSAC proceeds by repeatedly sampling randomly selected subsets of the data, each containing the minimal number of correspondences required to compute the relation exactly. In the case of  $F$  these minimal subsets contain seven correspondences. Each subset yields either one or three hypotheses for the parameters of  $F$  (see [6] for details).

The level of support for each hypothesis is evaluated using a robust consensus score applied to the whole data set. In our experiments each correspondence is scored using the robust cost function suggested by [12],

$$\begin{aligned} \rho(d^2) &= d^2 & , \text{ if } d^2 < T_{inlier} \\ &= T_{inlier} & , \text{ otherwise} \end{aligned} \quad (1)$$

, where  $d^2 = d_F^2(\mathbf{x}, \mathbf{x}')$  is the Sampson approximation to the squared  $F$ -manifold re-projection distance [6] for a correspondence  $c = (\mathbf{x}, \mathbf{x}')$  and fundamental matrix  $F$ .  $T_{inlier}$  is an inlier/outlier threshold based on the expected feature localization accuracy (see [6]). Sampling continues until a termination criterion is met (discussed below). The hypothesis with the best consensus score is the final estimate. Note that this score is really a **penalty** score, bounded below at zero, and monotonically increasing as additional correspondences are considered, hence **lower = better**.

From inspection of the algorithm in table 1 it is clear that the overall complexity of RANSAC is a product of the number of samples (and hence hypotheses) that are tested

and the average cost of scoring each hypothesis. The method proposed in this paper is aimed at reducing the latter. It is similar in spirit to Matas and Chum’s  $T_{d,d}$  [7] test, which attempts to avoid scoring the entire set of correspondences by first applying a  $d$ -out-of- $d$  pre-test to a randomly selected subset of size  $d$ . Only if *all*  $d$  are found to be inliers does the full scoring proceed. The authors suggest a value of  $d = 1$  is a most effective choice.

<p><b>Prerequisites</b></p> <ul style="list-style-type: none"> <li>• A set of feature correspondences <math>\mathcal{C}</math></li> <li>• A means of sampling minimal basis sets <math>S</math> from <math>\mathcal{C}</math></li> <li>• A means of generating model parameter hypotheses <math>p</math> given a basis <math>S</math></li> <li>• A robust correspondence scoring function <math>\rho(p, c)</math></li> </ul> <p><b>Algorithm</b></p> <p><math>n := 0</math>  <math>n_{max} := \infty</math>  <b>While</b> <math>n &lt; n_{max}</math></p> <p style="padding-left: 20px;">(1) Sample a basis set <math>S</math> of size <math>m</math> from <math>\mathcal{C}</math></p> <p style="padding-left: 20px;">(2) Generate a set of model hypotheses <math>\mathcal{P}</math> from <math>S</math></p> <p style="padding-left: 20px;">For each hypothesis <math>p</math> in <math>\mathcal{P}</math></p> <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> <p style="padding-left: 20px;"><i>If using <math>T_{d,d}</math> test</i></p> <p style="padding-left: 40px;">If <math>\rho(p, c) &gt; T_{inlier}</math> for any of <math>d</math> randomly chosen <math>c \in \mathcal{C}</math>, continue to next hypothesis.</p> </div> <p style="padding-left: 20px;">(3) Evaluate consensus score <math>C_p = \sum_{c \in \mathcal{C}} \rho(p, c)</math> and set of inliers <math>\mathcal{I}</math></p> <p style="padding-left: 20px;">(4) If <math>C_p &lt; C_{best}</math></p> <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> <p style="padding-left: 20px;"><i>If using LO-RANSAC</i></p> <p style="padding-left: 40px;">Refine model parameters <math>p</math> using all inliers <math>\mathcal{I}</math></p> <p style="padding-left: 40px;">Re-evaluate consensus score <math>C_p</math>, as step (3)</p> </div> <p style="padding-left: 40px;"><math>C_{best} := C_p, p_{best} := p</math></p> <p style="padding-left: 20px;">(5) Recompute <math>n_{max} = \frac{\log \eta}{\log(1 - \epsilon^m)}</math>, where <math>\epsilon = \frac{ \mathcal{I} }{ \mathcal{C} }</math> (eqn. 3).</p> <p style="padding-left: 20px;"><math>n := n + 1</math></p> <p><b>Return</b> <math>p_{best}</math>.</p>
---

Table 1: The vanilla RANSAC algorithm comprises steps 1 to 5. Matas & Chum’s  $T_{d,d}$  pre-evaluation test, and Chum *et al.*’s LO-RANSAC modification are also shown boxed.

Before proceeding to discuss the proposed method, it is necessary briefly to consider the details of the global termination criterion that indicates when enough samples have been taken. The standard termination criterion for RANSAC (see [10]) applies the following reasoning. Given the true fraction of inlying correspondences  $\epsilon$ , the probability of selecting a basis set of size  $m$  that consists entirely of inliers is  $\epsilon^m$ . Hence the probability

of sampling  $k$  basis sets all of which are polluted by at least one outlier is

$$\eta = (1 - \varepsilon^m)^k \quad (2)$$

Therefore the minimum number of samples  $k$  that must be taken in order that this probability falls below a given confidence threshold  $\eta^*$  is given by

$$k \geq \frac{\log \eta^*}{\log(1 - \varepsilon^m)} \quad (3)$$

Of course, the true inlier fraction  $\varepsilon$  is typically unknown before RANSAC begins. However, every time a new best hypothesis is discovered, a lower bound  $\hat{\varepsilon} \leq \varepsilon$  may be computed, and hence the number of samples  $k$  that are required in order to be confident that we have sampled at least one unpolluted basis may be updated as RANSAC progresses.

Matas and Chum’s  $T_{d,d}$  test requires a small but significant modification to this criterion. In this case, there is an additional probability  $(1 - \varepsilon^d)$  that a good hypothesis will fail the pre-evaluation test. Hence the expression for the minimal number of samples required becomes

$$k \geq \frac{\log \eta^*}{\log(1 - \varepsilon^{(m+d)})} \quad (4)$$

implying that use of the  $T_{d,d}$  test requires that more samples be drawn before termination than with “vanilla” RANSAC. This modification turns out to be very important, as we shall see in section 5. It is noted by Tordoff and Murray [9] and by Chum *et al.* [2] that these termination criteria can be overly optimistic, especially when the correspondences themselves are rather noisy, since in such cases a basis sample that contains only inliers to the correct model may not itself yield a model hypothesis with high support. However, in the experiments performed in this paper, the standard termination criterion, with confidence threshold  $\eta^* = 0.01$ , appears to be sufficient to ensure that all the methods tested return similar and satisfactory solutions for the epipolar geometry.

### 3 The hypergeometric bail-out test (HG-test)

In this section we describe the principal contribution of this paper : a novel early bail-out test that provides significant computational savings in the computation of the robust consensus score for a given model hypothesis (step 3 in table 1).

**A trivial early bail test** Before describing our proposed scheme, it is worth mentioning an extremely trivial early bail test that is not often noted in the literature, but which can (and should) always be used. It works as follows. The consensus scoring function is typically implemented as a simple loop over all the correspondences  $\mathcal{C}$ , during which the accumulated score  $C_i = \sum_{i < |\mathcal{C}|} \rho(p, c_i)$  grows monotonically. Remember that  $C_i$  is really a *penalty* score, **better** hypotheses having **lower** scores. Hence, we can always bail-out early if, during the evaluation loop, a point is reached at which the score for the current hypothesis,  $C_i$ , exceeds (is worse than) the current  $C_{best}$ , since  $C_i$  can only increase (get worse) from this point on.

**A novel early bail test** The proposed bail-out test can be explained intuitively as follows. Given a hypothesis to be scored, we evaluate a *randomly selected* subset of size  $n$  of the correspondences and observe an inlier fraction  $\varepsilon_n$ . If  $\varepsilon_n$  is very much smaller than the current best  $\varepsilon_{\text{best}}$  then it's highly unlikely that evaluating the remaining correspondences will yield an inlier fraction greater than  $\varepsilon_{\text{best}}$ , so we need not continue.

More formally, considering a subset of correspondences  $\mathcal{C}_n$  of size  $n$ , the number of inliers  $\kappa_n$  contained within  $\mathcal{C}_n$  follows a hypergeometric distribution  $\kappa_n \sim \text{HypG}(\kappa, n, \bar{\kappa}, N)$ , where  $\bar{\kappa}$  is the total inlier count for the given hypothesis, and  $N = |\mathcal{C}|$  is the total number of correspondences. Imagine we have a current best hypothesis  $S_{\text{best}}$  with inlier count  $\bar{\kappa}_{\text{best}}$  and that we have partially evaluated some new hypothesis  $S$ , having so far considered a subset of  $n$  correspondences. Ideally, we would like to take a Bayesian stand-point and ask the question ‘‘Given the  $\kappa_n$  inliers that we’ve observed so far, what’s the probability that the total  $\bar{\kappa} > \bar{\kappa}_{\text{best}}$ ?’’

$$P(\bar{\kappa} > \bar{\kappa}_{\text{best}}) = \sum_{\bar{\kappa}=\bar{\kappa}_{\text{best}}}^N P(\bar{\kappa}|\kappa_n, n, N) \quad (5)$$

If this probability is below a given threshold, we can safely abort further scoring of  $S$ .

Although it is straightforward to calculate this probability, it is computationally rather expensive and, unfortunately, does not lend itself easily to tabulation or approximation. So instead, we consider a related, but much simpler, one-sided confidence test. Denoting the cumulative density function as  $HG(\kappa_0, n, \bar{\kappa}, N)$  (the probability that  $\kappa \leq \kappa_0$ ), a confidence  $P_{\text{conf}}$  lower bound  $\kappa_{\text{min}}$  on the number of inliers  $\kappa_n$  observed within  $\mathcal{C}_n$  is defined by

$$HG(\kappa_{\text{min}}, n, \bar{\kappa}, N) \leq P_{\text{conf}} < HG(\kappa_{\text{min}} + 1, n, \bar{\kappa}, N) \quad (6)$$

Our rejection hypothesis is that the total  $\bar{\kappa}$  is  $< \bar{\kappa}_{\text{best}}$ , so we set  $\bar{\kappa} = \bar{\kappa}_{\text{best}}$  in the above. If  $\kappa_n < \kappa_{\text{min}}$ , the rejection hypothesis is accepted and we can bail out.

Given the above, the proposed algorithm for scoring a hypothesis proceeds as follows. It is essential that correspondences are scored in **randomized order**. After each correspondence is tested, the number of inliers  $\kappa$  observed so far is compared to a lower bound  $\hat{\kappa}_{\text{min}}$  that is a function of the current best inlier count  $\hat{\kappa}$ , the size of evaluation set  $n = |\mathcal{C}_n|$  so far considered, and the total size  $N$ . If  $\kappa < \hat{\kappa}_{\text{min}}$  the evaluation loop is aborted without considering the remainder of the correspondences.  $P_{\text{conf}} = 0.01$  in all of our experiments. We shall refer to this test as the **HG-test**.

**Implementation details** The correspondences may be randomly pre-shuffled prior to starting RANSAC, which can be done efficiently in  $O(N)$  time. For small values of  $n$ , the hypergeometric lower bound  $\kappa_{\text{min}}(n, \hat{\kappa}, N)$  may be approximated by the binomial distribution

$$\kappa \sim \text{Bin}\left(n, \frac{\hat{\kappa}}{N}\right) = \text{Bin}(n, \hat{\varepsilon}) \quad (7)$$

for which values may be tabulated for the chosen value of  $P_{\text{conf}}$ . For large  $n$ , the hypergeometric distribution may be approximated by a normal distribution

$$\kappa \sim N(\mu, \sigma^2) = N\left(n\hat{\varepsilon}, n\hat{\varepsilon}(1-\hat{\varepsilon})\left(\frac{N-n}{N-1}\right)\right) \quad (8)$$

in which case the value  $z_{\text{conf}}$  of the  $P_{\text{conf}}$  lower confidence limit for a *unit* normal distribution may be computed/looked-up in advance. Then the value of  $\hat{\kappa}_{\text{min}}$  for particular values of  $n$  and  $\varepsilon$  is given by

$$\hat{\kappa}_{\text{min}} = \lfloor n\hat{\varepsilon} - z_{\text{conf}}\sigma \rfloor \quad (9)$$

The steps in the proposed hypothesis scoring function are shown in table 2.

```

r := 1 , k := 0 , C := 0
Apply a random shuffle to C
While r ≤ |C|
  ρr = ρ(p, C\{r})
  If ρr < Tinlier then k := k + 1
  C := C + ρr
  Compute kmin using eqn. 9
  If k < kmin then abort evaluation
  r := r + 1
Return C.

```

Table 2: The hypothesis scoring function using the hypergeometric bail-out test (HG-test). This function corresponds to step 3 in table 1.

## 4 Experiments

In this section we evaluate the effectiveness of the proposed test as compared to the  $T_{d,d}$  test and the LO-RANSAC variant. We also assess the performance of the proposed test *in combination* with LO-RANSAC.

The test data consists of ten 100 frame video sequences (1000 frames total) captured with a hand-held PAL DV camera and containing a mixture of indoor and outdoor scenes (see figure 1). Robust F estimation is performed between all consecutive frames of video as follows. Between 1200 and 1500 point-features are detected in each  $720 \times 576$  images. Putative point-to-point matches are found using a correlation-based similarity score and subject to a maximum disparity limit of  $\pm 10\%$  of the image width ( $\pm 72$  pixels). This yields between 645 and 1081 (mean=850) putative correspondences. Six RANSAC variants are then applied to estimate F : vanilla RANSAC, RANSAC with  $T_{d,d}$  test, LO-RANSAC, LO-RANSAC with  $T_{d,d}$  test, RANSAC with the HG-test, and LO-RANSAC with the HG-test.

The shots used were chosen to be quite challenging for geometry estimation. The indoor clips are shot in low-light with the subject close-up, so the video is rather noisy and the motion jerky and blurry. The outdoor shots feature complex structure (e.g. foliage) and non-rigid motion (e.g. foaming water). Consequently, estimation of the epipolar geometry in these images tends to be more difficult than one might expect, with rather low inlier fractions ranging from 0.37 to 0.88 (mean=0.60).

In all cases, the value of  $d$  in the  $T_{d,d}$  test is  $d = 1$  as suggested by [2]. The RANSAC global termination criterion is computed with  $\eta^* = 0.01$  (eqns. 3 and 4). The  $P_{\text{conf}}$  lower bound on the number of inliers  $\kappa$  in the hypergeometric bail-out test is  $P_{\text{conf}} = 0.01$  (i.e.  $< 1\%$  chance that  $\kappa < \hat{\kappa}_{\text{min}}$ ). The trivial bail-out test is employed as a baseline criterion in *all* variants except those labelled “No bail test”.

## 5 Discussion

It is clear from table 3 that both the  $T_{d,d}$  test and the proposed HG-test significantly reduce the required number of evaluations of the robust scoring function  $\rho(p, c)$ , with the hypergeometric+LO-RANSAC variant performing best (a factor of two fewer evaluations than with the  $T_{d,d}$  test). The average time required per image pair for each variant is shown in table 4. This indicates that the reduction in function evaluations achieved by the HG-test results in a significant reduction in the overall computational cost (a factor of two compared to the next best). Also note that the trivial bail-out test yields significant performance advantages compared to the “no bail test” variants.

It is interesting to note that those variants employing only the trivial bail-out test are not greatly slower than those employing the  $T_{d,d}$  test in these experiments. This is because a large part of the total computational burden in RANSAC is the actual generation of hypotheses from sample sets, which in the case of F requires construction and SVD/QR decomposition of a  $7 \times 9$  matrix and solution of a cubic polynomial. This takes around  $35\mu\text{s}$  per F hypothesis in our implementation. As shown in table 5, the  $T_{d,d}$  test requires many more hypotheses to be evaluated before its modified termination criterion (eqn. 4) is met. This tends to offset the efficiency gains made in consensus scoring.

By comparison, the HG-test variants manage to keep both the average cost of scoring each hypothesis *and* the total number of hypotheses tested to a minimum. The hypergeometric+LO-RANSAC method out-performs all other variants in this respect.

Finally, table 6 shows the average and minimum inlier fractions computed across each shot for each of the six RANSAC variants that were tested. The fact that there is very little variation across the columns indicates that each method is returning similar and satisfactory results for the epipolar geometry. This reassures us that each method is drawing a sufficient number of samples and that none is guilty of sacrificing “diligence” for speed.

<i>avg. n_eval</i> / $10^5$	Vanilla RANSAC	LO-RANSAC
No bail test	8.20	7.19
Trivial bail test	5.40	4.72
$T_{d,d}$ test	1.31	1.19
HG-test	0.55	0.51

Table 3: The average number of F Sampson distance evaluations per image pair that was required by each of the six methods over the course of the 1000 RANSAC runs. The variants using the HG-test perform significantly fewer evaluations than the  $T_{d,d}$  test, and both are a great improvement over vanilla RANSAC.

<i>avg. time</i> / msec	Vanilla RANSAC	LO-RANSAC
No bail test	166	149
Trivial bail test	113	102
$T_{d,d}$ test	89.8	84.7
HG-test	42.0	40.3

Table 4: The average time taken per pair of images. Variants using the HG-test are significantly faster.

<i>avg. n.hypotheses</i>	Vanilla RANSAC	LO-RANSAC
No bail test	1026	898
Trivial bail test	1005	904
$T_{d,d}$ test	2139	1962
HG-test	1000	902

Table 5: The average number of putative F hypotheses that were required per pair of images. Compared to the vanilla and hypergeometric methods, the  $T_{d,d}$  method requires many more hypothesis samples to be generated before the global termination criterion (eqn. 4) is met. This significantly increases the computational overhead associated with using the  $T_{d,d}$  test.

Shot	Std	$T_{d,d}$	LO	LO+ $T_{d,d}$	HG-test	HG-test+LO
0	0.51 (0.41)	0.52 (0.41)	0.52 (0.42)	0.52 (0.41)	0.52 (0.41)	0.52 (0.42)
1	0.47 (0.37)	0.47 (0.38)	0.48 (0.38)	0.48 (0.38)	0.47 (0.37)	0.48 (0.38)
2	0.49 (0.42)	0.5 (0.44)	0.5 (0.43)	0.5 (0.44)	0.5 (0.42)	0.5 (0.43)
3	0.65 (0.54)	0.65 (0.55)	0.66 (0.54)	0.65 (0.55)	0.65 (0.55)	0.66 (0.54)
4	0.65 (0.54)	0.65 (0.55)	0.66 (0.54)	0.65 (0.56)	0.65 (0.55)	0.66 (0.54)
5	0.82 (0.74)	0.83 (0.75)	0.83 (0.75)	0.83 (0.76)	0.82 (0.74)	0.83 (0.75)
6	0.56 (0.47)	0.56 (0.48)	0.56 (0.47)	0.57 (0.48)	0.56 (0.47)	0.56 (0.47)
7	0.65 (0.42)	0.65 (0.42)	0.65 (0.43)	0.65 (0.42)	0.65 (0.42)	0.65 (0.43)
8	0.54 (0.46)	0.54 (0.46)	0.55 (0.47)	0.55 (0.46)	0.54 (0.46)	0.55 (0.47)
9	0.68 (0.43)	0.68 (0.41)	0.68 (0.39)	0.68 (0.44)	0.68 (0.43)	0.68 (0.39)

Table 6: The average and minimum (in brackets) inlier fraction in each of the ten 100 frame test shots at termination of each of six methods. Within each shot (row), there is very little variation across the columns, indicating that all of the methods are returning sufficiently similar results for the epipolar geometry, and none are guilty of sacrificing “diligence” for speed.

## 6 Conclusions

In this paper, we have proposed a novel and simple bail-out criterion that increases the efficiency of RANSAC by reducing the average computation cost associated with the consensus scoring of individual hypotheses. The proposed test has been evaluated for the robust estimation of the fundamental matrix using a large data set, and has been shown to provide significant computational savings compared to recently proposed RANSAC variants. The proposed approach has two key advantages. First, it can always be applied and will provide efficiency gains no matter how large or small the set of correspondences. Second, the bigger the set of correspondences, the greater are the benefits of effective early bail-out.

Future work should evaluate the effectiveness of the approach for the estimation of other multi-view relations, such as homographies and the trifocal tensor. The approach should also be evaluated for problems involving very large data sets ( $> 10^4$  correspondences).



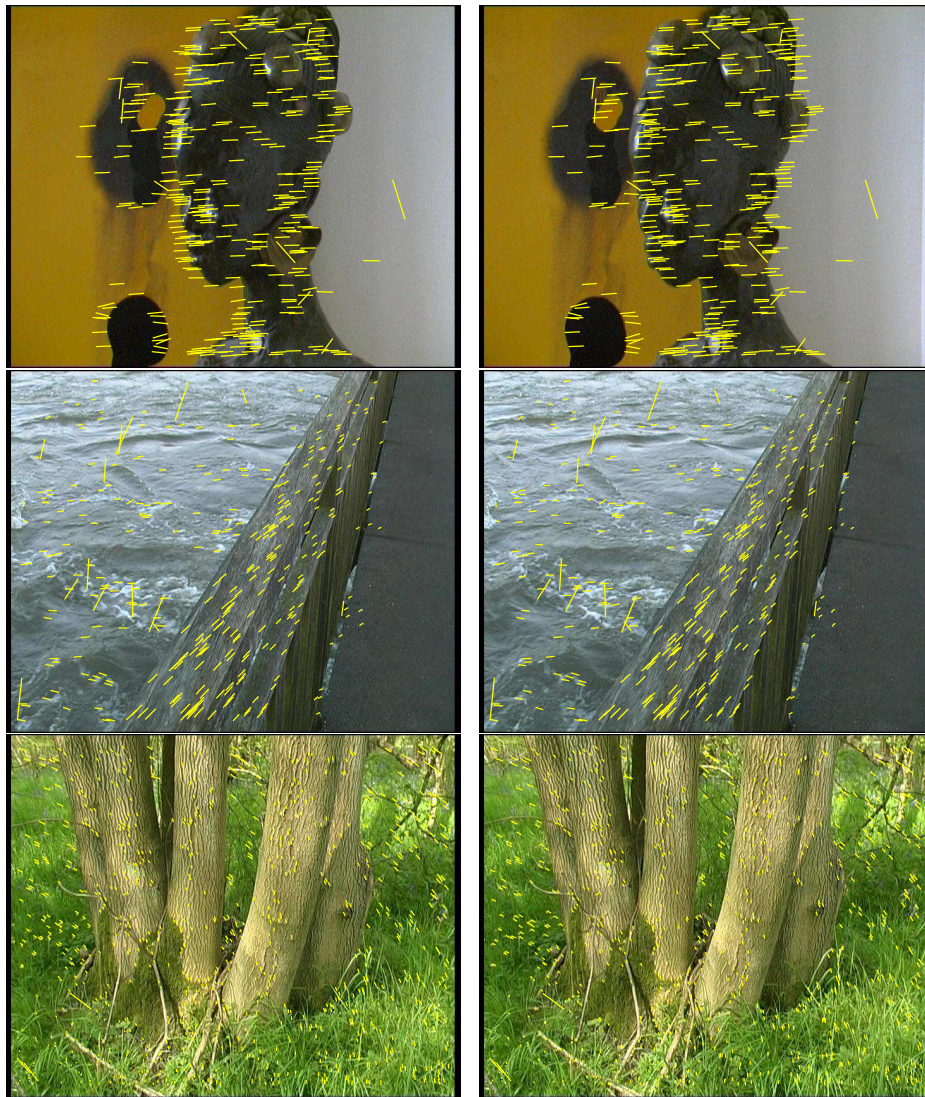


Figure 1: Three pairs of images from the evaluation set of ten 100 frame shots (1000 frames total). The video was captured with a hand-held PAL DV camera. The indoor clips are shot in low-light with the subject close-up, so the video is rather noisy and the motion jerky. The outdoor shots feature complex structure (e.g. foliage) and non-rigid motion (e.g. foaming water). Consequently, the feature correspondence problem for these images is actually quite challenging. Inliers to the epipolar geometry are super-imposed in yellow.

## References

- [1] O. Chum, J. Matas, and J.V. Kittler. Locally optimized RANSAC. In *Proc. German Association for Pattern Recognition (DAGM)*, pages 236–243, 2003.
- [2] O. Chum, J. Matas, and S. Obdrzalek. Enhancing RANSAC by generalized model optimization. In *Proc. Asian Conference on Computer Vision*, 2004.
- [3] O. Chum, T. Werner, and J. Matas. Epipolar geometry estimation via RANSAC benefits from the oriented epipolar constraint. In *Proc. International Conference on Pattern Recognition*, pages I: 112–115, 2004.
- [4] O. Chum, T. Werner, and T. Pajdla. Joint orientation of epipoles. In *Proc. British Machine Vision Conference*, 2003.
- [5] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. Assoc. Comp. Mach.*, 24(6):381–395, 1981.
- [6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 1998.
- [7] J. Matas and O. Chum. Randomized RANSAC with T(d,d) test. In *Proc. British Machine Vision Conference*, page Computer Vision Tools, 2002.
- [8] D. Nister. Preemptive RANSAC for live structure and motion estimation. In *Proc. International Conference on Computer Vision*, pages 199–206, 2003.
- [9] B. Tordoff and D.W. Murray. Guided sampling and consensus for motion estimation. In *Proc. European Conference on Computer Vision*, pages 82–98, 2002.
- [10] P. H. S. Torr. *Motion segmentation and outlier detection*. PhD thesis, Dept. of Engineering Science, University of Oxford, 1995.
- [11] P. H. S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–605, 1997.
- [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] T. Werner and T. Pajdla. Chirality in epipolar geometry. In *Proc. International Conference on Computer Vision*, pages I: 548–553, 2001.