ESTIMATING THE ESSENTIAL MATRIX: GOODSAC VERSUS RANSAC

Eckart Michaelsen[a], Wolfgang von Hansen[a], Michael Kirchhof[a], Jochen Meidow[a], Uwe Stilla[b]

[a]FGAN-FOM, Gutleuthausstr. 1, 76275 Ettlingen, Germany, mich@fom.fgan.de
[b]Photogrammetry and Remote Sensing, Technische Universität München, Arcisstr. 21, 80280 München, Germany

KEY WORDS: essential matrix, robust estimation, RANSAC, structure from motion

ABSTRACT:

GOODSAC is a paradigm for estimation of model parameters given measurements that are contaminated by outliers. Thus, it is an alternative to the well known RANSAC strategy. GOODSAC's search for a proper set of inliers does not only maximize the sheer size of this set, but also takes other assessments for the utility into account. Assessments can be used on many levels of the process to control the search and foster precision and proper utilization of the computational resources. This contribution discusses and compares the two methods. In particular, the estimation of essential matrices is used as example. The comparison is performed on synthetic and real data and is based on standard statistical methods, where GOODSAC achieves higher precision than RANSAC.

1 INTRODUCTION

1.1 Motivation

One of the basic tasks for many computer vision applications is to describe a set of measurements by a mathematical model. Often this model is overdetermined because the number of measurements is much higher than the number of unknown sought model parameters. Different methods to find an optimal solution despite the presence of noise and outliers have been devised during the years. The techniques used for robust estimation include random sampling (RANSAC), a complete search to test all possible inlier sets, clustering such as the Hough transform and maximum-likelihood-type estimation (McGlone et al., 2004, p. 103 ff). Even though they vary greatly in detail, their common property is to reduce the influence of outliers on initial solutions thus allowing their detection and removal.

RANSAC is often used when the outlier rate is high. The approach to its solution is to detect an inlier set of maximal size, while the precision of the resulting parameters is not taken into account. Better results can be achieved, when some of the often redundant inlier samples are traded in for others that have more impact on the parameters.

For real-time applications the computation time constraints are hard. Methods where the overall processing time is data-dependent and not guaranteed to be within a given bound cannot be accepted. In particular, it is desirable to have a method that can be terminated by external requirements. On the other hand, it should utilize the given resources properly, i. e. it should also not be ready long before the requirement for a result arrives and keep the resources idling for the rest of the time. For the experiments in this contribution, methods are chosen that have this anytime capability.

1.2 Our approach

In this paper, we propose the GOODSAC *(good sample consensus)* paradigm as an alternative to RANSAC *(random sample consensus)* (Fischler and Bolles, 1981). RANSAC uses a blind generate-and-test strategy that treats all samples equally regardless of their quality and requires a large number of tests to find an optimal solution. Furthermore, RANSAC is nondeterministic so that two runs on the same dataset will return different results.

In contrast to this, GOODSAC replaces the random sampling with an assessment driven selection of *good* samples. Good samples are those that possess a high degree of confidence and advantageous geometry so that the model parameters computed are well defined. Appropriate utility functions can usually be derived from the mathematical, geometric model in order to produce a sorted list of samples which features the most promising ones in the leading positions. Multiple abstraction stages from the measurements to the samples evade the complete search through all combinatorial possibilities.

A second enhancement is the replacement of tedious inlier tests by a clustering in parameter space. While RANSAC defines the best result as the one which has the largest inlier set, GOODSAC looks for solutions that turn up often. The concept of the inlier set is still present in GOODSAC through the set of predecessors – those samples who lead to the cluster in parameter space.

This contribution focuses on the application of GOODSAC to the estimation of essential matrix constraints between two images and in particular on the special case of nonuniform distributed interest points. Its performance is evaluated by comparing it to RANSAC with respect to robustness and accuracy. The experimental setup is chosen to resemble a situation frequently encountered in forward looking aerial thermal videos.

1.3 Related work

RANSAC (Fischler and Bolles, 1981) had been introduced to the scientific community 25 years ago and is widely used for its robustness in the presence of many outliers (25 Years of RANSAC, 2006). Documented enhancements of RANSAC based estimation mainly focus on the reduction of the required number of random samples in order to decrease processing time. (Matas et al., 2002) replaced the seven point correspondences required for the estimation of a fundamental matrix with three matching affine regions. Many implementations contain additions to the original RANSAC – typically constraints that bail out early on bad random samples.

A major recent improvement on the method itself is the preemptive RANSAC scheme (Nistér, 2005). While the original algorithm generates and tests one hypothesis at a time, preemptive RANSAC first generates a number of hypotheses and then tests them in parallel, discarding bad hypotheses early in order to speed up processing.

Another variant more along the lines of our work is PROSAC *(progressive* RANSAC*)* (Chum and Matas, 2005). PROSAC introduces an assessment component to narrow the set of samples to draw from. If successful, this achieves a higher inlier rate so that fewer runs are needed. In contrast, GOODSAC eliminates the random sampling process completely and introduces additional assessment components to achieve robustness.

GOODSAC has already been used for completely different recognition tasks in the past (Michaelsen et al., 2006).

2 METHOD AND TEST SETUP

2.1 The GOODSAC Paradigm

The good sample consensus principle has been designed for tasks where an assessment on the quality of the samples and of their parts can be provided (Michaelsen and Stilla, 2003). In this case the random search for a sample that maximizes consensus can be replaced by a controlled search. It is intended to capture sensible heuristics or proven utilities in a more systematic way and use them to prevent waste of computational resources and foster precision.

Let $F(\mathbf{m}, \mathbf{x}) = 0$ denote an implicitly defined functional model, where $\mathbf{m}$ is a k-tuple of parameters and $\mathbf{x}$ is an n-tuple of measurements. Only a subset $\{x_i\}$ with $i \in \mathcal{I} \subseteq \{1, \ldots, n\}$ fulfills the model. The task is to estimate $\mathbf{m}$ from $\mathbf{x}$. If the set $\mathcal{I}$ were known, the solution would be found by minimizing $\sum_{i \in \mathcal{I}} F(\mathbf{m}, x_i)^2$ under variation of $\mathbf{m}$. Systematical complete search in the power set 2^n for an optimal set $\mathcal{I}$ is usually not feasible. It is assumed that there exists $\ell < n$ minimally such that $\mathbf{m}$ can be determined from an ℓ-sample $\{x_{i_1}, \ldots, x_{i_\ell}\}$.

Whereas the RANSAC method draws samples at random, GOODSAC regards them as objects. Each object is assessed according to its presumed utility for the estimation task at hand. Furthermore, it exploits part-of hierarchies: intermediate objects are introduced between the single measurements and the ℓ-samples, which are smaller sub-sample objects (e. g. pairs or triples). Thus, the way is open for a better control on the search. Badly composed sub-sample objects can be neglected, while presumably well suited ones can be preferred – ℓ-sample objects vote for specific settings of the parameters $\mathbf{m}$ of the model. ℓ-sample objects with consistent votes – according to a metric and threshold in $\mathcal{M} \ni \mathbf{m}$ – are parts of a cluster object, which represents the highest level of the object hierarchy. The best cluster object is the result.

2.2 Assessment Driven Control

GOODSAC uses a general control approach. A part-of hierarchy is formulated as finite production system:

$$\big\{p_\iota; p_\iota = o_\kappa \leftarrow (o_\lambda, o_\mu) \vee o_\kappa \leftarrow \{o_\lambda, \ldots, o_\lambda\}\big\}, \qquad (1)$$

where p_ι denotes the productions and o_κ, o_λ or o_μ respectively denote object types. Note that the productions may be of two forms – either a more complex object is formed from an ordered pair of simpler objects of possibly different kinds or it is formed from a set of objects of the same kind. Associated with each production p_ι there is a predicate π_ι that the right side must fulfill for the production to be appropriate, a function that determines the object instances attribute values on the left side from the values found in the right side and in particular an assessment function α_ι for the newly constructed object instance. A proper assessment driven control cycle for production systems has already been given by (Stilla et al., 1995):

1. Form working elements $(\alpha_0, o_\lambda, \xi, \text{nil})$ from a given set of primitive object instances, where ξ is a pointer to the object instance and α_0 its initial assessment.
2. Sort the set of working elements according to the assessments α.
3. Pick a fixed number of elements from the "good end" of the sorted working set and proceed with all of them.
4. Let $(\alpha_0, o_\lambda, \xi, \chi)$:
   (a) If $\chi = \text{nil}$, then clone $(\alpha_0, o_\lambda, \xi, \chi)$ with $\chi = \iota$ for each production p_ι in which o_λ occurs on the right side.
   (b) Else query the database for partner object instances that fulfill π_ι together with the object o_λ to which ξ points; generate all new objects o_κ that are obtained using these combinations and insert new working elements $(\alpha_\iota, o_\kappa, \eta, \text{nil})$ for each of these with η pointing to them.
5. If the set of working elements is still not empty and no external break criterion holds continue at step 2.

After breaking the control cycle the best object of the highest hierarchical type is chosen as result. Using this control scheme for GOODSAC is achieved by taking the measurements $\mathbf{x}$ as primitive objects and larger samples as intermediate non-primitive objects up to the minimal ℓ-samples required for calculating F. Thus these objects can be attributed with parameter estimations for F. A single cluster production for these minimal ℓ-sample objects is added (of the type $o_\kappa \leftarrow \{o_\lambda, \ldots, o_\lambda\}$) that demands adjacency in the parameter space of F. It constructs non-minimal sample objects that are the result of the process.

The assessment functions used in this process must not only be capable of comparing the presumed utility of objects of the same type and hierarchy level, they must also be valid between objects of all different types, because all these objects constantly compete for the same computational resources. There are no random choices in a GOODSAC search run. It is completely determined by F, $\mathbf{x}$, the object hierarchy and the assessment functions. Its success depends on the care of the designer of the latter structures. It is particularly appropriate where utility assessment criteria can be given in a mathematically sound way. Sect. 2.3 gives an example for the estimation of essential matrix constraints.

2.3 An Example: Estimating Essential Matrices

GOODSAC is particularly suitable for essential matrix estimation. The minimal sample for this task is five correspondences $(x, y, x', y')^\top$ giving one constraint each, so that the five parameters of an essential matrix $\mathbf{E}$ can be obtained by evaluating the roots of a polynomial of 10th degree (Nistér, 2004). GOODSAC clusters each of the up to ten hypotheses computed from one sample. This independent treatment of the hypotheses is similar to a straightforward RANSAC implementation.

The hierarchy of the corresponding GOODSAC system consists of five object types: Correspondences $\boldsymbol{K}$, pairs $\boldsymbol{P}$, quadruples $\boldsymbol{Q}$, quintuples $\boldsymbol{R}$ and essential matrix clusters $\boldsymbol{C}$. The following attributes and assessment functions are assigned to these objects:

1. Objects $\boldsymbol{K}$ are obtained from image pairs taken from a video stream. Therefore they are attributed with the locations of the corresponding item in the first and second image

$(x, y, x', y')^\top$. It will be most useful if objects $\boldsymbol{K}$ with high assessment have a small expected error or outlier probability. There are extraction methods that provide these measures, and they have proven very useful for RANSAC acceleration (Chum and Matas, 2005). In our comparison uniform distributed random assessments have been used in the experiments in sect. 3 because emphasis is on exploitation of the geometric configurations. Random assessments are the worst possible choice apart from using the sequence in which the objects have been generated.

2. Objects $\boldsymbol{P} \leftarrow \boldsymbol{KK}$ are pairs of correspondence pairs. A similarity transform with four degrees of freedom may be calculated from an object $\boldsymbol{P}$. The expected deviation of this transform from the correct value would inversely depend on the Euclidean distance d between the objects $\boldsymbol{K}$ preceding it (in one or the other image). This motivates heuristically that the assessment is obtained from this distance. It is directly and linearly transformed to the assessment interval $[0, 1]$ by $a(\boldsymbol{P}) = d/d_{\max}$ where $d_{\max}$ is the maximal possible distance in the image.

3. Objects $\boldsymbol{Q} \leftarrow \boldsymbol{PP}$ are quadruple of correspondence objects. For their assessment the area a of the smallest of the four triangles formed from the four locations is used. This motivation is based on the fact that from an object $\boldsymbol{Q}$, a planar projective transform with eight degrees of freedom may be calculated. The expected deviation of this transform from the correct value would be highly correlated to this assessment value $a/a_{\max}$, where $a_{\max}$ is the area of the largest possible triangle in the image bounds. It can at most be equal to one, but usually it is much smaller. In order to balance between different object types, $(a/a_{\max})^e$ is used for the assessments of objects $\boldsymbol{Q}$ with an appropriate value $0 < e < 1$.

4. Objects $\boldsymbol{R} \leftarrow \boldsymbol{QK}$ are a quintuple of correspondence objects. Given an object $\boldsymbol{Q}$, partners $\boldsymbol{K}$ that are not co-linear with two of the four locations of this object are searched. The assessment is again obtained from the area of the smallest triangle. Also this assessment must be properly normed to be bounded by one and balanced with the other assessments as described above.

5. Objects $\boldsymbol{C} \leftarrow \{\boldsymbol{R}; \text{mutually consistent}\}$ are clusters of essential matrices very similar to each other. All objects $\boldsymbol{K}$ preceding them are again entered into the same procedure – this time in its overdetermined version – resulting in a new more precise estimation of $\mathbf{E}$. For assessment, the convex hull of all preceding objects $\boldsymbol{K}$ in the image is determined. The assessment value is formed as product of the number of correspondences k and the area of the convex hull. This assessment is neither balanced with respect to the other assessments nor bounded by one. It is not used for control purposes. Objects $\boldsymbol{C}$ do not compete for computational resources. This assessment is only used for picking the best result after termination of the GOODSAC search run.

If a very precise result is needed, a concluding consensus set may be formed from all objects $\boldsymbol{K}$ being consistent with the result, followed by least squares optimization. This last step is also proposed for RANSAC search. It is well known as "guided matching" (Hartley and Zisserman, 2000, p. 125).

2.4 Performance Evaluation

In this section we evaluate the performance of the proposed estimations with respect to the robustness of the procedures and the accuracies of the individual results. Whereas the ability of detecting outliers is specified by an error rate in terms of a binary classification, the results of the parameter estimations will be evaluated using statistical tests and results from adjustment theory, cf. (Mikhail, 1976, Förstner, 1994) for instance. For the real data set we do not have the true projection matrices and we are also not sure of the presence of possible non-projective distortions. Therefore we will describe qualitatively the result on a particular example.

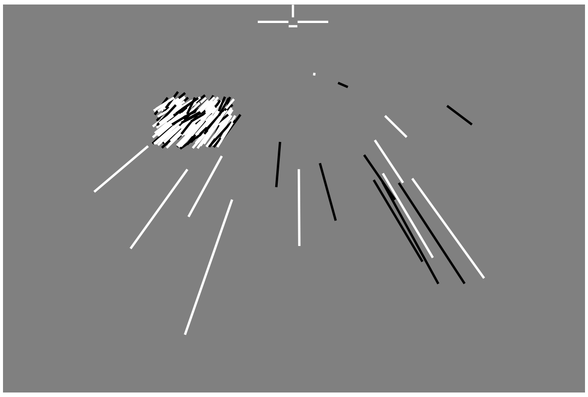

Figure 1: White displacements are inlier correspondences, black displacements outliers, the white aircraft symbol indicates the epipole.

2.4.1 Robustness

Outlier detection: Error rate. We consider the procedures to be a binary classifier indicating inliers and outliers with the help of a threshold. Competing classifiers can be evaluated based on their empirical confusion matrices. The rate of missed outlier detections is of interest beside the error rate being the ultimate measure of the classification performance (Jain et al., 2000). Since these measures are random variables, they have an associated distribution permitting hypothesis testing.

Self-diagnosis. In automatic analysis there is a demand for reliable self-diagnostics. Concerning the detectability of errors evaluation quantities can be derived from the stochastic model within general least squares adjustment models:

An initial covariance matrix $\mathbf{Q}_{xx}$ of the observations $\mathbf{x}$ is assumed to be known and related to the true covariance matrix $\mathbf{C}_{xx}$ by $\mathbf{C}_{xx} = \sigma_0^2 \mathbf{Q}_{xx}$ with the possibly unknown scale factor σ_0^2, also called variance factor. If the initial covariance matrix correctly reflects the uncertainties of the observations, this factor is $\sigma_0^2 = 1$. The estimated parameters are independent with respect to scaling of this covariance matrix, therefore only the ratios of the variances and covariances have to be known in advance.

The variance factor can be estimated from the estimated corrections $\hat{\mathbf{v}}$ for the observations $\mathbf{x}$ via

$$\hat{\sigma}_0^2 = \frac{\hat{\mathbf{v}}^\top \mathbf{Q}_{xx}^{-1} \hat{\mathbf{v}}}{R} \tag{2}$$

with the redundancy R of the system.

If the mathematical model actually holds and the observations are normally distributed, the estimated variance factor will be Fisher distributed with R and ∞ degrees of freedom (McGlone et al., 2004)

$$T_1 = \frac{\hat{\sigma}_0^2}{\sigma_0^2} \sim F_{R,\infty} \tag{3}$$

with the test statistic T_1 having the expectation value one.

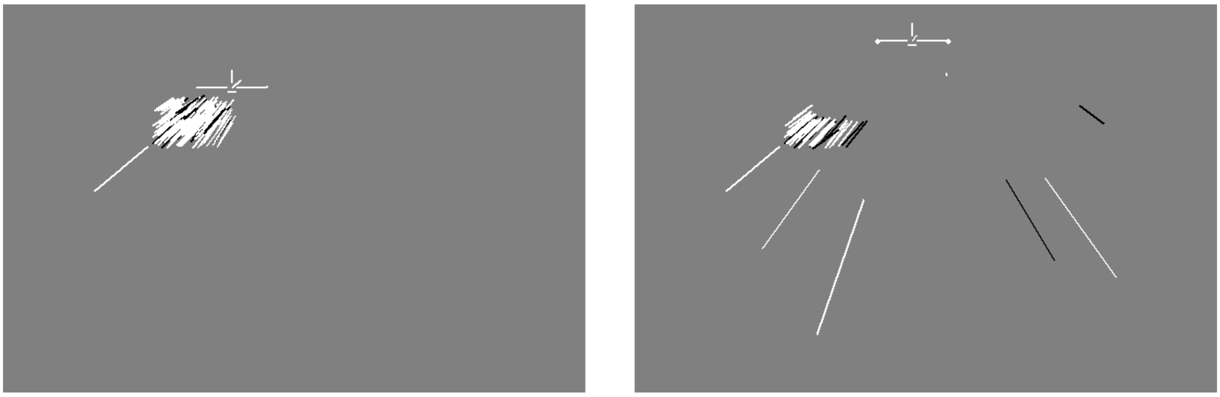

Figure 2: Left: Typical RANSAC result, right: Typical GOODSAC result. RANSAC maximizes the size of the inlier set, while GOODSAC returns correspondences that allow better precision of the estimated parameters.

If the test (3) is accepted, the data and model will fit. In the case of deviations there is no evidence for the reasons. In particular, small errors in the assumptions concerning the precision of the observations lead to a rejection of the test. But, for *synthetic data* σ_0^2 is known and the mathematical model holds. Therefore this test checks indirectly the robustness, since gross errors or blunders lead to a rejection.

2.4.2 Precision and Accuracy

Acceptability of the empirical precision. The empirical precision indicates the effect of random errors onto the estimated parameters, taking the estimated variance factor σ_0^2 into account. The empirically estimated covariance matrix for the estimated parameters $\mathbf{m}$ is

$$\hat{\mathbf{C}}_{\hat{m}\hat{m}} = \hat{\sigma}_0^2 \mathbf{Q}_{\hat{m}\hat{m}} \quad (4)$$

where the covariance matrix of the estimated parameters results from $\mathbf{Q}_{\hat{m}\hat{m}} = (\mathbf{J}^\top \mathbf{Q}_{xx}^{-1} \mathbf{J})^{-1}$ if the observations can be expressed explicit in terms of the parameters, where $\mathbf{J}$ denotes the Jacobian of the equations with respect to the parameters.

If a certain precision of the parameters is required, the individual values can be compared with some pre-specified tolerances for the specific application.

Empirical Accuracy. The evaluation of the covariance matrices, as discussed so far, is only an internal evaluation relying on the internal redundancy of the observation process. Systematic errors, which may not have an influence on the residuals but may deteriorate the estimated parameters, are not taken into account. Evaluating the empirical accuracy of the estimated parameters therefore requires reference values $\mathbf{m}_r$ for the parameters.

The Mahalanobis distance is useful for checking the complete set

$$T_2 = (\hat{\mathbf{m}} - \mathbf{m}_r)^\top \left(\mathbf{C}_{rr} + \hat{\mathbf{C}}_{\hat{m}\hat{m}}\right)^{-1} (\hat{\mathbf{m}} - \mathbf{m}_r) \sim \chi_u^2 \quad (5)$$

within a combined statistical test with u degrees of freedom being the number of parameters. If the test (5) has been rejected it can be concluded that the accuracy potential of the observations is not exploited, provided that the reference data $\mathbf{m}_r$ actually are correct and thus $\mathbf{m} \sim N(\mathbf{m}_r, \mathbf{C}_{rr})$.

3 EXPERIMENTS

3.1 Experiments with Synthetic Data

Optimal statistical analysis can only be accomplished with synthetic data because exact measure of the noise is required. A time constraint has been introduced by limiting the number of sample objects $\boldsymbol{R}$ to 2,000. The same number of samples was then permitted to the RANSAC search runs. This is almost two orders of magnitude larger than the standard textbook literature recommends for quintuple samples at 95% probability for an inlier-only sample with our 33% outlier rate (Hartley and Zisserman, 2000). For correspondences uniform distributed all over the image both methods will yield robust results. In order to elaborate the difference in the behavior a critical situation was simulated in the following way: 90% of the correspondences are located within a small region of the image. Only 10% are uniform distributed over the entire image (Fig. 1). For the synthetic data the scene is assumed to be flat. Therefore the correspondences result from a planar projective homography constraint. While a planar scene cannot be used for fundamental matrix estimation it should not pose a problem to essential matrix estimation following (Nistér, 2004).

Fig. 1 shows the generated frame-to-frame point correspondences. The camera motion has no rotational component. Thus the epipole – sketched as aircraft symbol – is a fixed point giving the flight direction and the horizon is a straight line of fixed points. 67% of the data are disturbed by an additive normally distributed shift error on the position in the second image. 33% of the data are disturbed by a much larger additive shift error on the position in the second image. This error is distributed uniformly within a squared search window eight times larger than the standard deviation of the inliers. GOODSAC requires assessment values for the correspondences. In this example, they were chosen randomly and independent of the outlier property and displacement error. Because the result of the search depends on these initial assessments the outcome is nondeterministic (as it normally would be with real assessments), allowing a statistical evaluation. Therefore the GOODSAC run was repeated 20 times with independently drawn assessments.

One GOODSAC result on the particular data set given in Fig. 1 is displayed in Fig. 2, left. The GOODSAC estimation is based on a fairly small number of objects $\boldsymbol{K}$ with some outliers included. These correspondences, however, are well spread over the image, so that the resulting estimation fits the ground truth neatly. The epipole is again displayed as aircraft symbol. Almost no rotation is left. Yaw and pitch rotations are indicated by a line inside the center of the aircraft symbol showing the resulting displacement and the roll is indicated by fins on the wingtips. RANSAC is a non-deterministic method. Therefore we repeated the experiment 20 times for each particular setting of correspondences. Among these runs there were also examples, where the outcome was superior to the GOODSAC result. To make our point clearly we decided to show an example result in Fig. 2 which comes up with

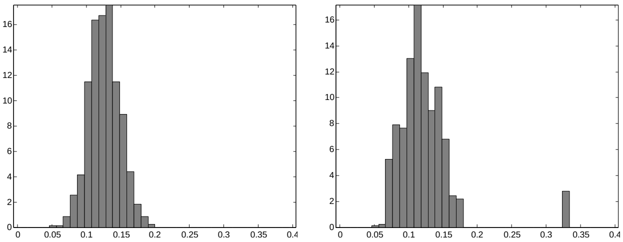

Figure 3: Empirical distributions for the false positives rates. Left: RANSAC, right: GOODSAC. For the peak at 33% see text.

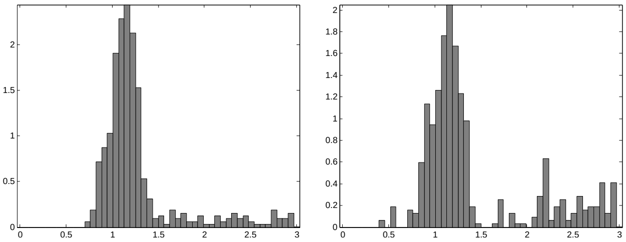

Figure 4: Empirical distribution of the variance ratios (3). Left: RANSAC, right: GOODSAC.

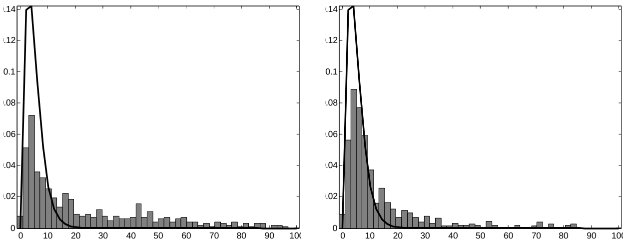

Figure 5: Empirical distribution of the Mahalanobis distances with the χ^2_5 distribution. Left: RANSAC, right: GOODSAC.

a fairly low false positive rate but with an unpleasant deviation of the essential matrix. There are only few black lines visible. But the estimation is based on a small image region. Thus the epipole can be displaced considerably from the true position. To compensate for this a considerable rotation – even in roll angle – is "invented".

The quantitative evaluation is based on 40 different settings of correspondences with 20 GOODSAC searches and 20 RANSAC searches performed on each. RANSAC always finds the vast majority of the consensus set inside densely populated areas. GOODSAC typically spreads the hypothesis generating samples across the entire image. Fig. 3 shows the distributions of the false positives rates for the GOODSAC and the RANSAC approach with a similar shape. The expectation value of 0.12 is caused by the fact that even in the optimal case about 30% of the outliers fit to the essential matrix and therefore are not detectable at this stage. There are rare situations, where almost all correspondence objects $\boldsymbol{K}$ closest to the image margin are actually outliers. This may cause the GOODSAC search to fail completely. Even after 2,000 quintuple objects $\boldsymbol{R}$ have been constructed no cluster may be found at all. Then the procedure falls back on using all correspondences as inliers. These cases lead to a small peak at 33% false positive rate for the GOODSAC method.

The empirical distribution of test statistics (Eq. (3)) is plotted in Fig. 4. Note that these values stem from different Fisher distributions since the degrees of freedom are varying with the number of inliers. The values obviously do not exceed the expectation value one for both estimation methods. Thus, the outliers have been removed successfully.

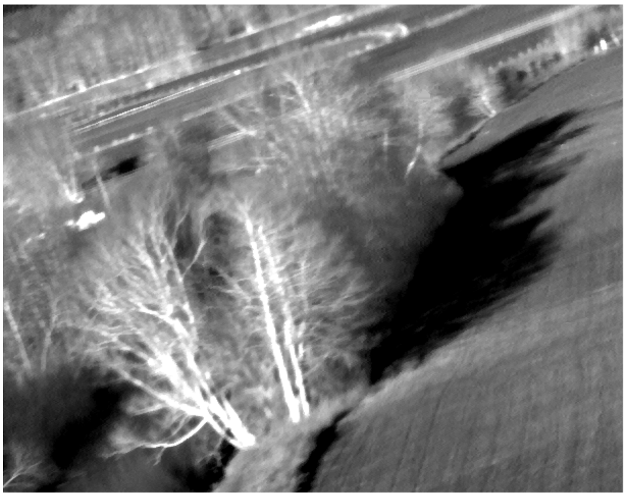

Figure 6: Frame from a forward-looking thermal video captured from a helicopter.

The empirical distributions of the Mahalanobis distances (5) shown in Fig. 5 reveal some deviation from the expected (analytical) probability density function. It can be seen that GOODSAC is closer to the theoretical distribution than RANSAC. This is because for a given sample and a corresponding essential matrix, the essential matrix is extrapolated outside the convex hull of this sample. While RANSAC maximizes the size of the sample as expected, this extrapolation leads to lower accuracy in the essential matrix.

3.2 An Experiment with Real Data

GOODSAC has been designed for applications where non-uniform distributed features are a common phenomenon. In particular, aerial forward-looking thermal videos often exhibit large uniform areas and strongly textured or structured regions often are quite sparse and small. An example frame is presented in Fig. 6.

Correspondences were obtained from a pair of frames with a sufficient baseline length, so that the displacements allow essential matrix estimation. Then GOODSAC and RANSAC were applied to these data under the same conditions that were also used for the synthetic setup. Quantitative evaluation of this experiment would need manual labeling of outliers, acquisition of ground truth, e. g. by an inertial navigation system, and repetition with a considerable number of image pairs. This has not yet been undertaken. Instead, in this contribution only the tendency of the outcome can be shown by example results in Fig. 7.

Note that while the sample found as consensus set by RANSAC has a larger size than the consensus set found by GOODSAC. However, some correspondences on the margin of the correspondence point cloud are missing in the RANSAC set, but appear in the set found by GOODSAC. This confirms the tendency found by the investigations with the synthetic data.

4 DISCUSSION AND OUTLOOK

Concerning the false positive rates on the synthetic dataset GOODSAC is only a little better than RANSAC. However, the Mahalanobis distance plots of the essential matrices resulting from the same experiments indicate that higher accuracy can be expected from GOODSAC. This can be explained by the fact that RANSAC simply tries to maximize the number of inliers which will only be directly related to the accuracy, if they are uniformly

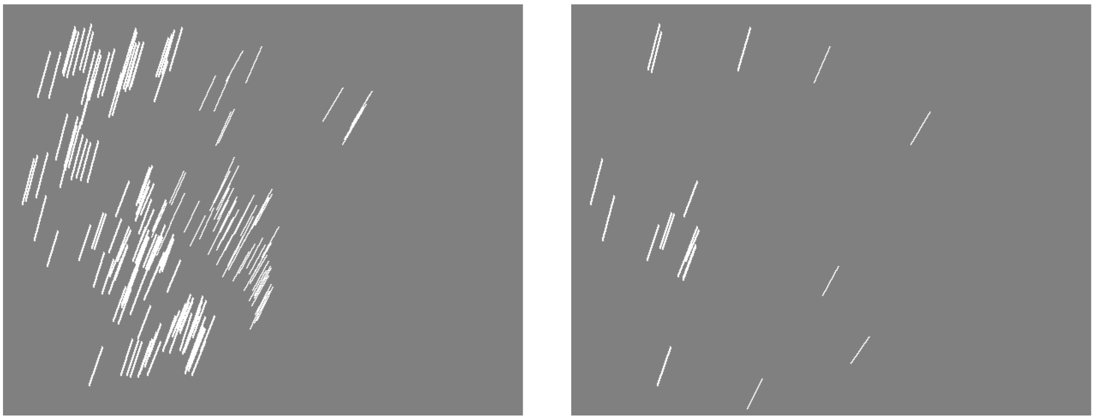

Figure 7: Typical result of the generated samples. Left: RANSAC, right: GOODSAC.

distributed over the entire image. GOODSAC tries to back the estimation by a stable geometric base and trades the sheer number of measurements for it.

The part-of hierarchy used for essential matrix estimation overlaps highly with that suitable for planar homography estimation (Michaelsen and Stilla, 2003). We may just add attributes to the quadruple objects $\boldsymbol{Q}$, add a clustering production and balance the assessment functions accordingly. If the scene is planar, the homography results will usually be more reliable, else the essential matrix solution will be better, while both calculations may be based on the same partial sub-calculations.

An open research problem with respect to this multiple use of intermediate objects is the choice of the assessment functions. We did not use any meaningful assessments on the elementary correspondence objects $\boldsymbol{K}$ here – for reason of fair competition. But in a real application we would of course use something reasonable: The quality of the match between the first and second image gives a good criterion related to both the outlier probability and the expected displacement error of an inlier correspondence. Or the length of displacement between the two images, because this estimation will fail on a set of stationary correspondences.

Further research is also needed to compare the two methods on real data with real inliers and outliers. The used assumptions on which the distributions of both kinds of correspondences are based must be verified. Here, we presume the normal distribution of the inliers to be the smaller problem. The distribution of real inliers may be deviating due to the pixel structure of the detector or properties of the matching algorithm, but the deviation may well be tolerable. The distribution of outliers, however, is probably a more severe problem. Real outliers do not occur randomly with equal probability anywhere. They are caused by unpredictable clutter effects. Some of these (e. g. moving objects, partial occlusions) may be foreseeable but a quantitative prediction is hard. They will, however, have a bias, and the influence of outliers on either method remains to be studied. However, the presented statistics based on the simplified assumption on the outliers still indicate potential usefulness of the presented method.

REFERENCES

25 Years of RANSAC, 2006. Workshop in conjunction with CVPR 2006, New York, June 18.

Chum, O. and Matas, J., 2005. Matching with PROSAC – Progressive Sample Consensus. In: Proc. of Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 220–226.

Fischler, M. A. and Bolles, R. C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the Association for Computing Machinery 24(6), pp. 381–395.

Förstner, W., 1994. Diagnostics and Performance Evaluation in Computer Vision. In: Performance versus Methodology in Computer Vision, NSF/ARPA Workshop, IEEE Computer Society, Seattle, pp. 11–25.

Hartley, R. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge.

Jain, A. K., Duin, R. P. W. and Mao, J., 2000. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Recognition and Machine Intelligence 22(1), pp. 4–37.

Matas, J. et al., 2002. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In: Proceedings of the British Machine Vision Conference, Vol. 1, pp. 384–393.

McGlone, J. C., Mikhail, E. M. and Bethel, J. (eds), 2004. Manual of Photogrammetry. 5th edn, American Society of Photogrammetry and Remote Sensing.

Michaelsen, E. and Stilla, U., 2003. Good Sample Consensus Estimation of 2D-Homographies for Vehicle Movement Detection from Thermal Videos. In: H. Ebner, C. Heipke, H. Mayer and K. Pakzad (eds), Photogrammetric Image Analysis PIA03, International Archives of Photogrammetry and Remote Sensing, Vol. 34, Part 3/W8, pp. 125–130.

Michaelsen, E., Soergel, U. and Thoennessen, U., 2006. Perceptual Grouping in Automatic Detection of Man-Made Structure in High Resolution SAR Data. Pattern Recognition Letters 27(4), pp. 218–225.

Mikhail, E. M., 1976. Observations and Least Squares. With Contributions by F. Ackerman. University Press of America, Lanham.

Nistér, D., 2004. An Efficient Solution to the Five Point Relative Pose Problem. IEEE Transactions on Pattern Recognition and Machine Intelligence 26(6), pp. 756–769.

Nistér, D., 2005. Preemptive RANSAC for Live Structure and Motion Estimation. Machine Vision and Applications 16(5), pp. 321–329.

Stilla, U., Michaelsen, E. and Lütjen, K., 1995. Structural 3D-Analysis of Aerial Images with a Blackboard-based Production System. In: A. Grün and O. Kübler (eds), Automatic Extraction of Man-made Objects from Aerial and Space Images, Birkhäuser, Basel, pp. 53–62.