MAGSAC: Marginalizing Sample Consensus

Daniel Barath[12], Jiri Matas[1], and Jana Noskova[1]

[1] Centre for Machine Perception, Department of Cybernetics
Czech Technical University, Prague, Czech Republic
[2] Machine Perception Research Laboratory, MTA SZTAKI, Budapest, Hungary

barath.daniel@sztaki.mta.hu

Abstract

A method called, σ-consensus, is proposed to eliminate the need for a user-defined inlier-outlier threshold in RANSAC. Instead of estimating the noise σ, it is marginalized over a range of noise scales. The optimized model is obtained by weighted least-squares fitting where the weights come from the marginalization over σ of the point likelihoods of being inliers. A new quality function is proposed not requiring σ and, thus, a set of inliers to determine the model quality. Also, a new termination criterion for RANSAC is built on the proposed marginalization approach. Applying σ-consensus, MAGSAC is proposed with no need for a user-defined σ and improving the accuracy of robust estimation significantly. It is superior to the state-of-the-art in terms of geometric accuracy on publicly available real-world datasets for epipolar geometry ($\boldsymbol{F}$ and $\boldsymbol{E}$) and homography estimation. In addition, applying σ-consensus only once as a post-processing step to the RANSAC output always improved the model quality on a wide range of vision problems without noticeable deterioration in processing time, adding a few milliseconds.[1]

1. Introduction

The RANSAC (RANdom SAmple Consensus) algorithm proposed by Fischler and Bolles [5] in 1981 has become the most widely used robust estimator in computer vision. RANSAC and its variants have been successfully applied to a wide range of vision tasks, e.g. motion segmentation [25], short baseline stereo [25, 27], wide baseline stereo matching [18, 13, 14], detection of geometric primitives [21], image mosaicing [7], and to perform [28] or initialize multi-model fitting [10, 17]. In brief, the RANSAC approach repeatedly selects random subsets of the input point set and fits a model, e.g. a plane to three 3D points or a homography to four 2D point correspondences. Next,

the quality of the estimated model is measured, for instance by the size of its support, i.e. the number of inliers. Finally, the model with the highest quality, polished e.g. by least squares fitting on its inliers, is returned.

Since the publication of RANSAC, a number of modifications has been proposed. NAPSAC [16], PROSAC [1] and EVSAC [6] modify the sampling strategy to increase the probability of selecting an all-inlier sample early. NAPSAC assumes that the inliers are spatially coherent, PROSAC exploits an a priori predicted inlier probability of the points and EVSAC estimates a confidence in each of them. MLESAC [26] estimates the model quality by a maximum likelihood process with all its beneficial properties, albeit under certain assumptions about inlier and outlier distributions. In practice, MLESAC results are often superior to the inlier counting of plain RANSAC and they are less sensitive to the user-defined inlier-outlier threshold. In MSAC [24], the robust estimation is formulated as a process that estimates both the parameters of the data distribution and the quality of the model in terms of maximum a posteriori. timates the model quality by a maximum likelihood process with all its beneficial properties, albeit under certain assumptions about inlier and outlier distributions.

One of the highly attractive properties of RANSAC is its small number of control parameters. The termination is controlled by a manually set confidence value η and the sampling stops as soon as the probability of finding a model with higher support falls below $1-\eta$.[2] The setting of η is not problematic, the typical values are 0.95 or 0.99, depending on the required confidence in the solution.

The second, and most critical, parameter is the inlier noise scale σ that determines the inlier-outlier threshold $\tau(\sigma)$ which strongly influences the outcome of the procedure. In standard RANSAC and its variants, σ must be provided by the user which limits its fully automatic out-of-the-box use and requires the user to acquire knowledge about the problem at hand. In Fig. 1, the inlier residuals

---

[1]The source code is at https://github.com/danini/magsac

[2]Note that the probabilistic interpretation of η holds only for the standard $\{0, 1\}$ cost function.

are shown for four real datasets demonstrating that σ varies scene-by-scene and, thus, there is no single setting which can be used for all cases.

To reduce the dependency on this threshold, MIN-PRAN [22] assumes that the outliers are uniformly distributed and finds the model where the inliers are least likely to have occurred randomly. Moisan et al. [15] proposed a contrario RANSAC, to optimize each model by selecting the most likely noise scale.

As the *major contribution* of this paper, we propose an approach, σ-consensus, that eliminates the need for σ, the noise scale parameter. Instead of σ, only an upper limit is required. The final outcome is obtained by weighted least-squares fitting, where the weights are given for *marginalizing* over σ, using likelihood of the model given data and σ. Besides finessing the need for a precise scale parameter, the novel method, called MAGSAC, is more precise than previously published RANSACs. Also, we propose a post-processing step applying σ-consensus to the *so-far-the-best-model* without noticeable deterioration in processing time, i.e. at most a few milliseconds. In our experiments, the method *always improved* the input model (coming from RANSAC, MSAC or LO-RANSAC) on a wide range of problems. Thus we see no reason for not applying it after the robust estimation finished. As a *second contribution*, we define a new quality function for RANSAC. It measures the quality of a model without requiring σ and, therefore, a set of inliers to measure the model quality. Moreover, as a *third contribution*, due to not having a single inlier set and, thus, an inlier ratio, the standard termination criterion of RANSAC is marginalized over σ to be applicable to the proposed method.

2. Notation

In this paper, the input points are denoted as $\mathcal{P} = \{p \mid p \in \mathbb{R}^k, k \in \mathbb{N}_{>0}\}$, where k is the dimension, e.g. $k = 2$ for 2D points and $k = 4$ for point correspondences. The inlier set is $\mathcal{I} \subseteq \mathcal{P}$. The model to fit is represented by its parameter vector $\theta \in \Theta$, where $\Theta = \{\theta \mid \theta \in \mathbb{R}^d, d \in \mathbb{N}_{>0}\}$ is the manifold, for instance, of all possible 2D lines and d is dimension of the model, e.g. $d = 2$ for 2D lines (angle and offset). Fitting function $F : \mathcal{P}^* \to \Theta$ calculates the model parameters from $n \geq m$ points, where $\mathcal{P}^* = \exp \mathcal{P}$ is the power set of $\mathcal{P}$ and $m \in \mathbb{N}_{>0}$ is the minimum point number for fitting a model, e.g. $m = 2$ for 2D lines. Note that F is a combined function applying different estimators on the basis of the input set, for example, a minimal method if $n = m$ and least-squares fitting otherwise. Function $D : \Theta \times \mathcal{P} \to \mathbb{R}$ is the point-to-model residual function. Function $I : \mathcal{P}^* \times \Theta \times \mathbb{R} \to \mathcal{P}^*$ selects the inliers given model θ and threshold σ. For instance, if the original RANSAC approach is considered, $I_{\text{RANSAC}}(\theta, \sigma, \mathcal{P}) = \{p \in \mathcal{P} \mid D(\theta, p) < \sigma\}$, for truncated quadratic dis-

(a) `homogr` dataset   (b) EVD dataset
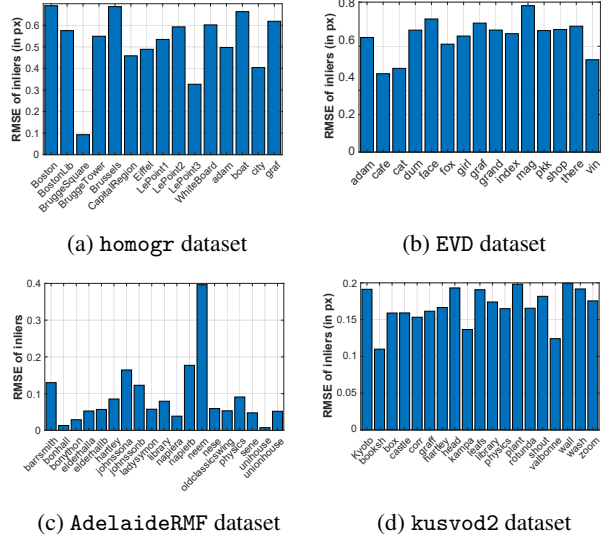
(c) `AdelaideRMF` dataset   (d) `kusvod2` dataset

Figure 1: The average residuals (RMSE in pixels; vertical axis) of manually annotated inliers given the ground truth model for each scene (horizontal) of four datasets.

Notation			
$\mathcal{P}$	- Set of data points	σ	- Noise standard deviation
θ	- Model parameters	D	- Residual function
I	- Inlier selector function	Q	- Model quality function
F	- Fitting function	m	- Minimal sample size
$\tau(\sigma)$	- Inlier-outlier threshold	$\sigma_{\max}$	- Upper bound of σ

tance of MSAC, $I_{\text{MSAC}}(\theta, \sigma, \mathcal{P}) = \{p \in \mathcal{P} \mid D^2(\theta, p) < 9/4\sigma^2\}$. The quality function is $Q : \mathcal{P}^* \times \Theta \times \mathbb{R} \to \mathbb{R}$. Higher quality is interpreted as better model. For RANSAC, $Q_{\text{RANSAC}}(\theta, \sigma, \mathcal{P}) = |I(\theta, \sigma, \mathcal{P})|$ and for MSAC, it is

$$Q_{\text{MSAC}}(\theta, \sigma, \mathcal{P}) = \sum_{i=1}^{|I(\theta, \sigma, \mathcal{P})|} \left(1 - \frac{D^2(\theta, I_i(\theta, \sigma, \mathcal{P}))}{9/4\sigma^2}\right),$$

where $I_i(\theta, \sigma, \mathcal{P})$ is the ith inlier.

3. Marginalizing sample consensus

A method called MAGSAC is proposed in this section eliminating the threshold parameter from RANSAC-like robust model estimation.

3.1. Marginalization over σ

Let us assume the noise σ to be a random variable with density function $f(\sigma)$ and let us define a new quality function for model θ marginalizing over σ as follows:

$$Q^*(\theta, \mathcal{P}) = \int Q(\theta, \sigma, \mathcal{P}) f(\sigma) \mathrm{d}\sigma. \qquad (1)$$

Having no prior information, we assume σ being uniformly distributed, $\sigma \sim \mathcal{U}(0, \sigma_{max})$. Thus

$$Q^*(\theta, \mathcal{P}) = \frac{1}{\sigma_{max}} \int_0^{\sigma_{max}} Q(\theta, \sigma, \mathcal{P}) d\sigma. \qquad (2)$$

For instance, using $Q(\theta, \sigma, \mathcal{P})$ of plain RANSAC, i.e. the number of inliers, where σ is the inlier-outlier threshold and $\{D(\theta, p_i)\}_{i=1}^{|\mathcal{P}|}$ are the distances to model θ such that $0 \leq D(\theta, p_1) < D(\theta, p_2) < .... < D(\theta, p_K) < \sigma_{max} < D(\theta, p_{K+1}) < ... < D(\theta, p_{|\mathcal{P}|})$ we get a quality function

$$Q^*(\theta, \mathcal{P}) = K - \frac{1}{\sigma_{max}} \sum_{k=1}^K D(\theta, p_k) = \sum_{k=1}^K \left(1 - \frac{D(\theta, p_k)}{\sigma_{max}}\right).$$

Assuming the distribution of inliers and outliers to be uniform (inlier $\sim \mathcal{U}(0, \sigma)$; outlier $\sim \mathcal{U}(0, l)$) and using log-likelihood of model θ as its quality function Q, we get

$$Q^*(\theta, \mathcal{P}) = K(\ln \frac{l}{\sigma_{max}} + 1)$$

$$- \frac{1}{\sigma_{max}} \sum_{k=1}^K D(\theta, p_k)(1 + \ln \frac{l}{D(\theta, p_k)}) - |\mathcal{P}| \ln l. \qquad (3)$$

Typically, the residuals of the inliers are calculated as the Eucledian-distance from the model in some ρ-dimensional space. In case of assuming errors of the distances along each axis of this ρ-dimensional space to be independent and normally distributed with the same variance σ^2, (residuals)$^2/\sigma^2$ have chi-squared distribution with ρ degrees of freedom. Therefore,

$$g(r \mid \sigma) = 2C(\rho)\sigma^{-\rho} \exp\left(-r^2/2\sigma^2\right) r^{\rho-1}$$

is a density of residuals of inliers with

$$C(\rho) = \frac{1}{2^{\rho/2}\Gamma(\rho/2)},$$

where

$$\Gamma(a) = \int_0^{+\infty} t^{a-1} \exp(-t) dt$$

for $a > 0$ is the gamma function.

In MAGSAC, the residuals of the inliers are described by a distribution with density $g(r \mid \sigma)$, and the outliers by a uniform one on the interval $[0, l]$. Note that, for images, l can be set to the image diagonal. The inlier-outlier threshold $\tau(\sigma)$ is set to the 0.95 or 0.99 quantile of the distribution with density $g(r \mid \sigma)$. Consequently, the likelihood of model θ given σ is

$$L(\theta, \mathcal{P} \mid \sigma) = \frac{1}{l^{|\mathcal{P}|-|\mathcal{I}(\sigma)|}}$$

$$\qquad (4)$$

$$\prod_{p \in \mathcal{I}(\sigma)} \left[2C(\rho)\sigma^{-\rho} D^{\rho-1}(\theta, p) \exp\left(\frac{-D^2(\theta, p)}{2\sigma^2}\right)\right].$$

MAGSAC, for a given σ, uses log-likelihood of model θ as its quality function as follows: $Q(\theta, \sigma, \mathcal{P}) = \ln L(\theta, \mathcal{P}|\sigma)$. Thus, the quality marginalized over σ is the following.

$$Q_{\mathrm{MAGSAC}}^*(\theta, \mathcal{P}) = \frac{1}{\sigma_{max}} \int_0^{\sigma_{max}} \ln L(\theta, \mathcal{P}|\sigma) d\sigma$$

$$\approx -|\mathcal{P}| \ln l + \frac{1}{\sigma_{max}} \sum_{i=1}^K [i(\ln 2C(\rho)l - \rho \ln \sigma_i) \qquad (5)$$

$$- \frac{R_i}{\sigma_i^2} + (\rho - 1) L r_i](\sigma_i - \sigma_{i-1}),$$

where $\{D(\theta, p_i)\}_{i=1}^{|\mathcal{P}|}$ are the distances to model θ, $\sigma_0 = 0$ and $0 \leq D(\theta, p_1) = \tau(\sigma_1) < D(\theta, p_2) = \tau(\sigma_2) < ... < D(\theta, p_K) = \tau(\sigma_K) < \tau(\sigma_{max}) < D(\theta, p_{K+1}) < ... < D(\theta, p_{|\mathcal{P}|})$, $R_i = \frac{1}{2} \sum_{j=1}^i D(\theta, p_j)^2$ and $Lr_i = \sum_{j=1}^i \ln D(\theta, p_j)$. As a consequence, the proposed new quality function Q_{MAGSAC}^* does not depend on a manually set noise level σ.

3.2. σ-consensus model fitting

Due to not having a set of inliers which could be used to polish the model obtained from a minimal sample, we propose to use weighted least-squares fitting where the weights are the point probabilities of being inliers.

Suppose that we are given model θ estimated from a minimal sample. Let $\theta_\sigma = F(I(\theta, \sigma, \mathcal{P}))$ be the model implied by the inlier set $I(\theta, \sigma, \mathcal{P})$ selected using $\tau(\sigma)$ around the input model θ. It can be seen from Eq. 4 that the likelihood of point $p \in \mathcal{P}$ being inlier given model θ_σ is

$$L(p \mid \theta_\sigma, \sigma) = 2C(\rho)\sigma^{-\rho} D^{\rho-1}(\theta_\sigma, p) \exp\left(\frac{-D^2(\theta_\sigma, p)}{2\sigma^2}\right).$$

For finding the likelihood of a point being an inlier marginalized over σ, the same approach is used as before:

$$L(p \mid \theta) \approx \frac{2C(\rho)}{\sigma_{max}} \sum_{i=1}^K (\sigma_i - \sigma_{i-1})$$

$$\qquad (6)$$

$$\sigma_i^{-\rho} D^{\rho-1}(\theta_{\sigma_i}, p) \exp\left(\frac{-D^2(\theta_{\sigma_i}, p)}{2\sigma_i^2}\right).$$

and the polished model $\theta_{\mathrm{MAGSAC}}^*$ is estimated using weighted least-squares, where the weight of point $p \in \mathcal{P}$ is $L(p \mid \theta)$.

3.3. Termination criterion

Not having an inlier set and, thus, at least a rough estimate of the inlier ratio, makes the standard termination criterion of RANSAC [8] inapplicable, which is as follows:

$$k(\theta, \sigma, \mathcal{P}) = \frac{\ln(1 - \eta)}{\ln\left(1 - \left(\frac{|I(\theta, \sigma, \mathcal{P})|}{|\mathcal{P}|}\right)^m\right)}, \qquad (7)$$

where k is the iteration number, η a manually set confidence in the results, m the size of the minimal sample needed for the estimation, and $|I(\theta, \sigma, \mathcal{P})|$ is the inlier number of the so-far-the-best model.

In order to determine k without using a particular σ, it is a straightforward choice to marginalize similarly to the model quality. It is as follows:

$$k^*(\mathcal{P}, \theta) = \frac{1}{\sigma_{max}} \int_0^{\sigma_{max}} k(\theta, \sigma, \mathcal{P}) d\sigma$$
$$\approx \frac{1}{\sigma_{max}} \sum_{i=1}^{K} \frac{(\sigma_i - \sigma_{i-1}) \ln(1-\eta)}{\ln\left(1 - \left(\frac{|I(\theta, \sigma_i, \mathcal{P})|}{|\mathcal{P}|}\right)^m\right)}. \quad (8)$$

Thus the number of iterations required for MAGSAC is calculated during the process and updated whenever a new so-far-the-best model is found, similarly as in RANSAC.

## 4. Algorithms using σ-consensus

In this section, we propose two algorithms applying σ-consensus. First, MAGSAC will be discussed incorporating the proposed marginalizing approach, weighted least-squares and termination criterion. Second, a post-processing step is proposed which is applicable to the output of every robust estimator. In the experiments, it *always improved* the input model without noticeable deterioration in the processing time, adding maximum a few milliseconds.

4.1. Speeding up the procedure

Since plain MAGSAC would apply least-squares fitting a number of times, the implied computational complexity would be fairly high. Therefore, we propose techniques for speeding up the procedure. In order to avoid unnecessary operations, we introduce a σ_{max} value and use only the σs smaller than σ_{max} in the optimization procedure. Thus, from $\sigma_1 < \sigma_2 < ... < \sigma_K < \sigma_{max} < \sigma_{K+1} < ... < \sigma_n$ only $\sigma_1, \sigma_2, ...,$ and σ_i are used. This σ_{max} can be set to a fairly big value, for example, 10 pixels. In the case when the results suggest that σ_{max} is too low, e.g. if the density mode of the residuals is close to σ_{max}, the computation can be repeated with a higher value.

Instead of calculating θ_{σ_i} for every σ_i, we divide the range of σs uniformly into d partitions. Thus the processed set of σs are the following: $\sigma_1 + (\sigma_{max} - \sigma_1)/d$, $\sigma_1 + 2(\sigma_{max} - \sigma_1)/d, ..., \sigma_1 + (d-1)(\sigma_{max} - \sigma_1)/d, \sigma_{max}$. By this simplification, the number of least-squares fittings drops to d from K, where $d \ll K$. In the experiments, d was set to 10.

Also, as it was proposed for USAC [19], there are several ways of skipping early the evaluation of models which do not have the chance of being better than the previous so-far-the-best. For this purpose, we apply SPRT [2] with a τ_{ref} threshold. Threshold τ_{ref} is not used in the model evaluation
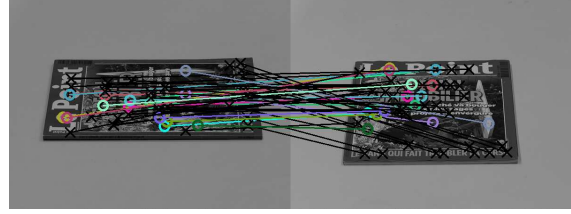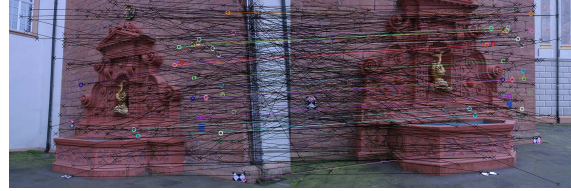

(a) Homography; `homogr` dataset. Errors: $\epsilon_{\text{LO-MSC}} = 4.3$ (2nd) and $\epsilon_{\text{MAGSAC}} = 2.9$ pixels (1st).
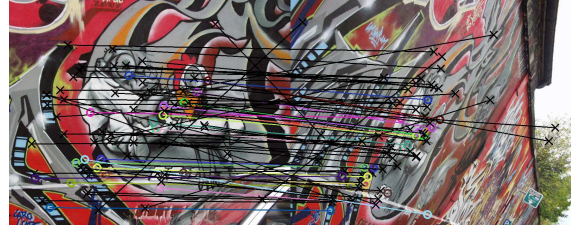
(b) Homography; `EVD` dataset. Errors: $\epsilon_{\text{LO-RSC}} = 9.1$ (2nd) and $\epsilon_{\text{MAGSAC}} = 4.4$ pixels (1st).

(c) Fundamental matrix; `kusvod2` dataset. Errors: $\epsilon_{\text{MSC}} = 14.3$ (2nd) and $\epsilon_{\text{MAGSAC}} = 0.5$ pixels (1st).

(d) Essential matrix; `Strecha` dataset. Errors: $\epsilon_{\text{MSC}} = 4.2$ (2nd) and $\epsilon_{\text{MAGSAC}} = 2.5$ pixels (1st).

(e) Essential matrix; `Strecha` dataset. Errors: $\epsilon_{\text{MSC}} = 5.6$ (2nd) and $\epsilon_{\text{MAGSAC}} = 3.9$ pixels (1st).

Figure 2: Example results of MAGSAC where it was significantly more accurate than the second most accurate method. Average errors (in pixels) are written in the captions. Inlier correspondences are drawn by color and outliers by black crosses.

or inlier selection steps, but is used merely to skip applying σ-consensus when it is unnecessary. In the experiments, τ_{ref}

was set to 1 pixel.

Finally, the parallel implementation of σ-consensus can be straightforwardly done on GPU or multiple CPUs evaluating each σ on a different thread. In our C++ implementation, it runs on multiple CPU cores.

4.2. The σ-consensus algorithm

The proposed σ-consensus is described in Alg. 1. The input parameters are: the data points ($\mathcal{P}$), initial model parameters (θ), a user-defined partition number (d), and a limit for σ ($\sigma_{\max}$).

As a first step, the algorithm takes the points which are closer to the initial model than $\tau(\sigma_{\max})$ (line 1). Function τ returns the threshold implied by the input σ parameter. In case of $\chi^2(4)$ distribution, it is $\tau(\sigma) = 3.64\sigma$. Then the residuals of the inliers are sorted, therefore, in $\{\sigma_i\}_{i=1}^{|\mathcal{I}|}$, $\sigma_i < \sigma_j \Leftrightarrow i < j$. In $\mathcal{I}_{\mathrm{ord}}$, the indices of the points are ordered reflecting to $\{\sigma_i\}_{i=1}^{|\mathcal{I}|}$, thus $\sigma_i = D(\theta, \mathcal{I}_{\mathrm{ord},i})/3.64$ (line 2). In lines 3 and 4, the weights are initialized to zero, and $\sigma_{\max}$ is set to $\max(\{\sigma_i\}_{i=1}^{|\mathcal{I}|})$. Then the current σ range is calculated. For instance, the first range to process is $[\sigma_1, \sigma_1 + \delta_\sigma]$. Note that $\sigma_1 = 0$ due to having at least m points at zero distance from the model. The cycle runs from the first to the last point and, since $\mathcal{I}_{\mathrm{ord}}$ is ordered, each subsequent point is farther from the model than the previous ones. Until the end of the current range, i.e. partition, is not reached (line 7), it collects the points (line 8) one-by-one. After exceeding the boundary of the current range, θ_σ is calculated using all the previously collected points (line 10). Then, for each point, the weight is updated by the implied probability (line 12). Finally, the algorithm jumps to the next range (line 13). After the weights have been calculated for each point, weighted least-squares fitting is applied to obtain the marginalized model parameters (line 14).

4.3. MAGSAC

The MAGSAC procedure polishing every estimated model by σ-consensus is shown in Alg. 2. First, it initializes the model quality to zero and the required iteration number to ∞ (line 1). In each iteration, it selects a minimal sample (line 3), fits a model to the selected points (line 4) validates it (line 5) and applies σ-consensus to obtain the parameters marginalized over σ (line 6). The validation step includes degeneracy testing and tests which stop the evaluation of the model if there is no chance of being better than the previous so-far-the-best, e.g. by SPRT test [2]. Note that, for SPRT, the validation step is also included into σ-consensus when the distances from the current model are calculated (line 1 in Alg. 1). Finally, the model quality is calculated (line 8), the so-far-the-best model and required iteration number are updated (line 10) if required (line 9). **As a post-processing step** in time sensitive applications, σ-consensus is a possible option for polishing the RANSAC output instead of

applying a least-squares fitting to the inliers. In this case, σ-consensus is applied only once, thus improving the results without noticeable deterioration in the processing time.

Algorithm 1 σ-consensus.

Input: $\mathcal{P}$ – points; θ – model parameters; d – partition number; $\sigma_{\max}$ – σ limit; η – confidence
Output: θ^* – optimal model parameters
1: $\mathcal{I} \leftarrow I(\mathcal{P}, \theta, \tau(\sigma_{\max}))$
2: $\mathcal{I}_{\mathrm{ord}}, \{\sigma_i\}_{i=1}^{|\mathcal{I}|} \leftarrow \mathrm{sort}(\{D(\theta, p)\}_{p \in \mathcal{I}})$
3: $\{w_i\}_{i=1}^{|\mathcal{I}|} \leftarrow \{0\}_{i=1}^{|\mathcal{I}|}, \sigma_{\max} \leftarrow \max(\{\sigma_i\}_{i=1}^{|\mathcal{I}|})$
4: $\delta_\sigma \leftarrow \sigma_{\max}/d, \sigma_{\mathrm{next}} \leftarrow \delta_\sigma, \mathcal{I}_{\mathrm{tmp}} \leftarrow \emptyset$
5: **for** i $= 1 \rightarrow |\mathcal{I}_{\mathrm{ord}}|$ **do**
6: $\quad p \leftarrow \mathcal{I}_{\mathrm{ord},i}, d_p \leftarrow D(\theta, p)$
7: $\quad$ **if** $d_p \leq \tau(\sigma_{\mathrm{next}})$ **then**
8: $\quad\quad \mathcal{I}_{\mathrm{tmp}} \leftarrow \mathcal{I}_{\mathrm{tmp}} \cup \{p\}$
9: $\quad\quad$ **continue**
10: $\quad \theta_\sigma \leftarrow F(\mathcal{I}_{\mathrm{tmp}})$
11: $\quad$ **for** i $= 1 \rightarrow |\mathcal{I}|$ **do**
12: $\quad\quad w_i \leftarrow w_i + W(\theta_\sigma, \mathcal{I}_i, \delta_\sigma)/\sigma_{\max}$ $\qquad \triangleright$ Eq. 6
13: $\quad \mathcal{I}_{\mathrm{tmp}} \leftarrow \mathcal{I}_{\mathrm{tmp}} \cup \{p\}, \sigma_{\mathrm{next}} \leftarrow \sigma_{\mathrm{next}} + \delta_\sigma$
14: $\theta^* \leftarrow F(\mathcal{I}, \{w_i\}_{i=1}^{|\mathcal{I}|})$ $\qquad \triangleright$ Weighted LSQ

Algorithm 2 MAGSAC

Input: $\mathcal{P}$ – data points; $\sigma_{\max}$ – σ limit; σ_{ref} – reference σ; m – sample size; d – partition number; η – confidence
Output: θ^* – optimal model; q^* – model quality
1: $q^* \leftarrow 0, k \leftarrow \infty$
2: **for** i $= 1 \rightarrow k$ **do**
3: $\quad \{p_j\}_{j=1}^m \leftarrow \mathrm{Sample}(\mathcal{P})$
4: $\quad \theta \leftarrow F(\{p_j\}_{j=1}^m)$
5: $\quad$ **if** $\neg\mathrm{Validate}(\theta, \sigma_{\mathrm{ref}})$ **then**
6: $\quad\quad$ **continue**
7: $\quad \theta' \leftarrow \sigma\text{-consensus}(\mathcal{P}, \theta, d, \sigma_{\max})$ $\qquad \triangleright$ Alg. 1
8: $\quad q' \leftarrow Q(\theta', \mathcal{P})$
9: $\quad$ **if** $q > q^*$ **then**
10: $\quad\quad q^*, \theta^*, k \leftarrow q', \theta', \mathrm{Iters}(q', |\mathcal{P}|, \eta)$ $\quad \triangleright$ Eq. 8

---

5. Experimental Results

To evaluate the proposed post-processing step, we tested several approaches with and without this step. The compared algorithms are: RANSAC, MSAC, LO-RANSAC, LO-MSAC, LO-RANSAAC [20], and a contrario RANSAC [15] (AC-RANSAC). LO-RANSAAC is a method including model averaging into robust estimation. AC-RANSAC estimates the noise σ. The same random seed was used for all methods and they performed a final least-squares on the obtained inlier set. The difference between RANSAC – MSAC and LO-RANSAC – LO-MSAC

is merely the quality function. Moreover, the methods with LO prefix run the local optimization step proposed by Chum et al. [3] with an inner RANSAC applied to the inliers. The parameters used are as follows: $\sigma = 0.3$ was the inlier-outlier threshold used for the RANSAC loop (this value was proposed in [11] and also suited for us). The number of inner RANSAC iterations was $r = 20$. The required confidence η was 0.95. There was a minimum number of iterations required (set to 20) before the first LO step applied and also before termination. The reported error values are the root mean square (RMS) errors. For σ-consensus, $\sigma_{\max}$ was set to 10 pixels for all problems. The partition of σ range was set to $d = 10$. Therefore, the processed set of σs were $\sigma_{\max}/d, 2\sigma_{\max}/d, ..., (d-1)\sigma_{\max}/d$, and $\sigma_{\max}$.

5.1. Synthesized Tests

To test the proposed method in a fully controlled environment, two cameras were generated by their 3×4 projection matrices $\mathbf{P}_1 = \mathbf{K}_1[\mathbf{I}_{3\times 3} \mid 0]$ and $\mathbf{P}_2 = \mathbf{K}_2[\mathbf{R}_2 \mid -\mathbf{R}_2\mathbf{t}_2]$. Camera $\mathbf{P}_1$ was located in the origin and its image plane was parallel to plane XY. The position of the second camera was at a random point inside a unit-sphere around the first one, thus $|\mathbf{t}_2| \leq 1$. Its orientation was determined by three random rotations affecting around the principal directions as follows: $\mathbf{R}_2 = \mathbf{R}_{X,\alpha}\mathbf{R}_{Y,\beta}\mathbf{R}_{Z,\gamma}$, where $\mathbf{R}_{X,\alpha}$, $\mathbf{R}_{Y,\beta}$ and $\mathbf{R}_{Z,\gamma}$ are 3D rotation matrices rotating around axes X, Y and Z, by α, β and γ degrees, respectively ($\alpha, \beta, \gamma \in [0, \pi/2]$). Both cameras had a common intrinsic camera matrix with focal length $f_x = f_y = 600$ and principal points $[300, 300]^\mathrm{T}$. A 3D plane was generated with random tangent directions and origin $[0, 0, 5]^\mathrm{T}$. It was sampled at n_i locations, thus generating n_i 3D points at most one unit far from the plane origin. These points were projected into the cameras. All of the random parameters were selected using uniform distribution. Zero-mean Gaussian-noise with σ standard deviation was added to the projected point coordinates. Finally, n_o outliers, i.e. uniformly distributed random point correspondences, were added. In total, 200 points were generated, therefore $n_i + n_o = 200$.

The mean results of 500 runs are reported in Fig. 3. The competitor algorithms are: RANSAC (RSC), MSAC (MSC), LO-RANSAC (LO-RSC), LO-MSAC (LO-MSC) and MAGSAC. Suffix "$+\sigma$" means that σ-consensus was applied as a post-processing step. Plots (a–c) reports the geometric accuracy (in pixels) as a function of the noise level σ using different outlier ratios (a – 0.2, b – 0.5, c – 0.8). The RANSAC confidence was set to 0.95. For instance, outlier ratio 0.8 means that $n_o = 160$ and $n_i = 40$. By looking at the differences between methods with and without the proposed post-processing step ("$+\sigma$"), it can be seen that it almost always improved the results. E.g. the geometric error of LO-MSC is higher than that of LO-MSC $+ \sigma$ for every noise σ. MAGSAC results are superior to

that of the competitor algorithms on every outlier ratio. It can be seen that it is less sensitive to noise and more robust to outliers. In (d), the processing time (in seconds) is reported as the function of the noise σ. MAGSAC is the slowest on the easy scenes, i.e. when the noise $\sigma < 0.3$ pixels. Thereafter, it becomes the fastest method due to requiring significantly fewer iterations than the others. Plots (e–f) of Fig. 3 demonstrate that the accuracy provided by MAGSAC cannot be achieved by simply letting RANSAC run longer. The charts report the results for a fixed iteration number, i.e. calculated from the ground truth inlier ratio and confidence set to 0.999. For outlier ratio 0.8, it was $\log(0.001)/\log(1 - 0.2^4) = 4\ 314$. For outlier ratio 0.9, it was $\log(0.001)/\log(1 - 0.1^4) = 69\ 074$. It can be seen that MAGSAC obtains significantly more accurate results than the competitor algorithms. It finds the desired model in most of the cases even when the outlier ratio is high.

5.2. Real World Experiments

In this section, MAGSAC and the proposed post-processing step is compared with state-of-the-art robust estimators on real-world data for fundamental matrix, homography and essential matrix fitting. See Fig. 2 for example image pairs where the error ($\epsilon_{\mathrm{MAGSAC}}$; in pixels) of the MAGSAC estimate was significantly lower than that of the second best method.

Fundamental Matrices. To evaluate the performance on fundamental matrix estimation we downloaded kusvod2[3] (24 pairs), Multi-H[4] (5 pairs), and AdelaideRMF[5] (19 pairs) datasets. Kusvod2 consists of 24 image pairs of different sizes with point correspondences and fundamental matrices estimated from manually selected inliers. AdelaideRMF and Multi-H consist a total of 24 image pairs with point correspondences, each assigned manually to a homography or the outlier class. All points which are assigned to a homography were considered as inliers and the others as outliers. In total, 48 image pairs were used from three publicly available datasets. All methods applied the seven-point method [8] as a minimal solver for estimating $\mathbf{F}$. Thus they drew minimal sets of size seven in each iteration. For the final least squares fitting, the normalized eight-point algorithm [9] was ran on the obtained inlier set. Note that all fundamental matrices were discarded for which the oriented epipolar constraint [4] did not hold.

The first three blocks of Table 1, each consisting of three rows, report the quality of the estimation on each dataset as the average of 100 runs on every image pair. The first two columns show the name of the tests and the investigated properties: **(1)** e_{avg} is the RMS geometric error in pixels of the obtained model w.r.t. the manually annotated in-

---

[3]http://cmp.felk.cvut.cz/data/geometry2view/
[4]http://web.eee.sztaki.hu/~dbarath/
[5]cs.adelaide.edu.au/~hwong/doku.php?id=data

(a) 20% outl., 95% conf. (b) 50% outl., 95% conf. (c) 80% outl., 95% conf.

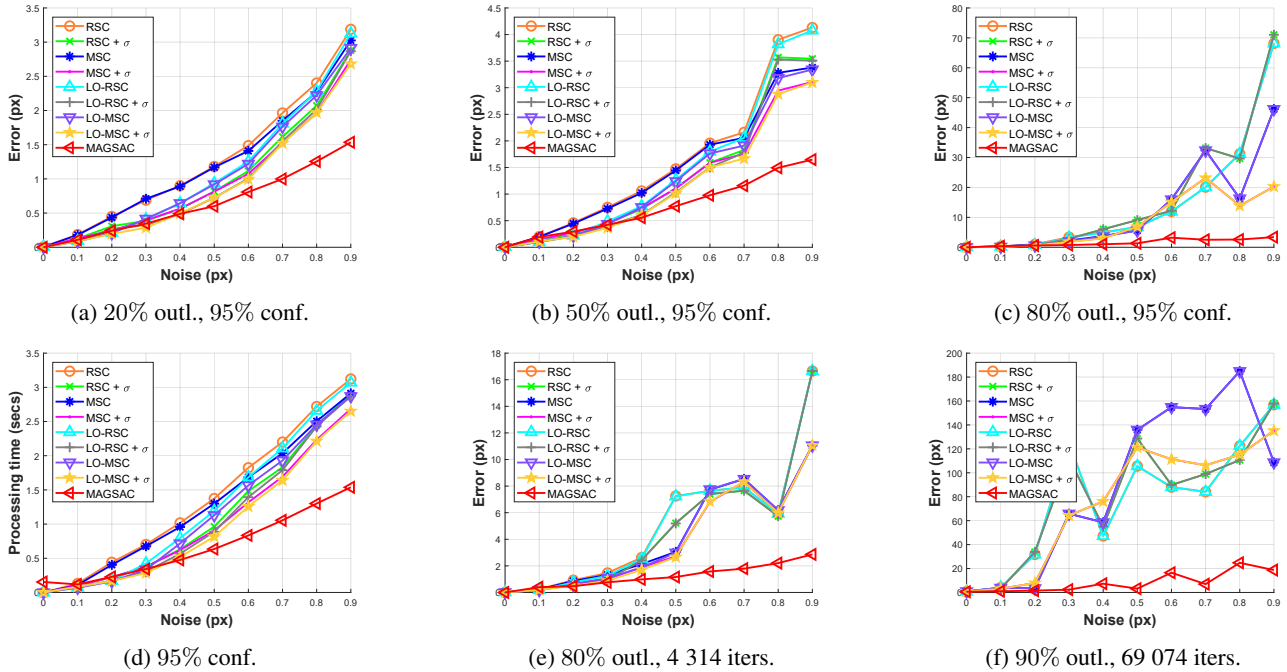(d) 95% conf. (e) 80% outl., 4 314 iters. (f) 90% outl., 69 074 iters.

Figure 3: *Synthetic homography fitting.* The competitor methods are: RANSAC, MSAC, LO-RANSAC, LO-MSAC and MAGSAC. Suffix "$+\sigma$" means that σ-consensus was applied to the output. Plots (a–c) report the errors (in pixels) as function of the noise σ with confidence set to 0.95. Plot (d) shows the avg. processing time (in seconds). Plots (e–f) report the results made by using a fixed iteration number calculated from the ground truth inlier ratio and confidence set to 0.999.

liers. For fundamental matrices and homographies, it is the average Sampson distance and re-projection error, respectively. For essential matrices, it is the mean Sampson distance of the implied **F** and the correspondences. **(2)** Value t is the mean processing time in milliseconds. **(3)** Value s is the mean number of samples, i.e. RANSAC iterations, had to be drawn till termination. Note that the iteration numbers of methods applied with or without the proposed post-processing are equal.

It can be seen that for **F** estimation the proposed post-processing step improved the results in nearly all of the tests with negligible deterioration in the processing time. The errors were reduced by approximately 8% compared with the methods without σ-consensus. MAGSAC led to the most accurate results for `kusvod2` and `Multi-H` datasets and it was the third best for `AdelaideRMF` dataset by a small margin of 0.03 pixels.

Homographies. To test homography estimation we downloaded `homogr` (16 pairs) and `EVD`[6] (15 pairs) datasets. Each consists of image pairs of different sizes from 329×278 up to 1712×1712 with point correspondences and inliers selected manually. The `Homogr` dataset consists of mostly short baseline stereo images, whilst the pairs of EVD undergo an extreme view change, i.e. wide baseline or ex-

treme zoom. All algorithms applied the normalized four-point algorithm [8] for homography estimation both in the model generation and local optimization steps. The 4th and 5th blocks of Fig. 1 show the mean results computed using all the image pairs of each dataset. Similarly as for **F** estimation, the proposed post-processing step always improved (by 1.42 pixels on average). For both datasets, the results obtained by MAGSAC were significantly more accurate than what the competitor algorithms obtained.

Essential Matrices. To estimate essential matrices, we used the `strecha` dataset [23] consisting of image sequences of buildings. All images are of size 3072×2048. The ground truth projection matrices are provided. The methods were applied to all possible image pairs in each sequence. The SIFT detector [12] was used to obtain correspondences. For each image pair, a reference point set with ground truth inliers was obtained by calculating **F** from the projection matrices [8]. Correspondences were considered as inliers if the symmetric epipolar distance was smaller than 1.0 pixel. All image pairs with less than 50 inliers found were discarded. In total, 467 image pairs were used in the evaluation. The results are reported in the 6th block of Table 1. The trend is similar to the previous cases. The most accurate essential matrices were obtained by MAGSAC. Also it was the fastest algorithm on average.

			RSC	$+\sigma$	MSC	$+\sigma$	LO-RSC	$+\sigma$	LO-MSC	$+\sigma$	LO-RSAAC	AC-RSC	**MAGSAC**
kusvod2	**F**, 24	e_{avg}	0.73	0.60	0.75	0.64	0.56	0.52	0.58	0.50	1.01	0.63	**0.38**
		t	38	39	19	19	25	25	**17**	**17**	**17**	55	31
		s	661	661	313	313	316	316	160	160	160	**71**	382
		fails	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	**0.00**
Adelaide	**F**, 19	e_{avg}	0.58	0.54	0.66	0.63	0.28	**0.27**	0.31	0.31	0.33	0.46	0.30
		t	491	493	420	420	393	394	**380**	**380**	**380**	447	939
		s	3 327	3 327	2 752	2 752	2 221	2 221	2 091	2 091	2 091	**2 047**	2 638
		fails	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Multi-H	**F**, 4	e_{avg}	0.70	0.59	0.84	0.75	0.53	0.52	0.50	0.50	0.58	0.72	**0.47**
		t	321	329	149	149	132	140	119	128	126	**46**	467
		s	1 987	1 987	908	908	580	580	327	327	327	**23**	1 324
		fails	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
homogr	**H**, 16	e_{avg}	3.61	2.12	3.64	2.18	3.39	2.13	3.53	2.19	2.95	1.83	**1.37**
		t	83	85	**64**	65	71	72	64	65	65	37	131
		s	1 815	1 815	1 395	1 395	1 478	1 478	1 222	1 222	1 222	**148**	877
		fails	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	**0.00**	0.06
EVD	**H**, 15	e_{avg}	5.73	4.08	5.15	3.57	5.42	4.07	4.78	3.55	4.55	5.05	**1.76**
		t	381	383	379	380	367	369	353	356	355	291	**162**
		s	6 212	6 212	6 106	6 106	5 847	5 847	5 540	5 540	5 540	3 463	**2 239**
		fails	0.57	0.50	0.57	0.43	0.57	0.50	0.57	0.43	0.53	0.33	**0.29**
strecha	**E**, 467	e_{avg}	7.05	6.91	7.32	7.13	9.61	9.48	10.62	10.23	10.17	15.56	**6.51**
		t	3 046	3 052	2 894	2 894	2 548	2 549	2 535	2 537	2 536	4 637	**2 398**
		s	3 530	3 530	3 315	3 315	2 789	2 789	2 770	2 770	2 770	3 680	**2 183**
		fails	0.24	0.22	0.26	0.22	0.27	0.22	0.26	0.22	0.24	0.23	**0.00**
all		e_{avg}	3.07	2.47	3.04	2.48	3.30	2.83	3.39	2.88	3.27	4.04	**1.80**
		e_{med}	2.17	1.36	2.24	1.47	1.98	1.33	2.06	1.35	1.98	1.28	**0.92**
		t	727	730	654	655	589	592	**578**	581	580	921	688
		fails	0.19	0.16	0.17	0.14	0.18	0.15	0.16	0.14	0.16	0.10	**0.03**

Table 1: *Accuracy of robust estimators on two-view geometric estimation.* Fundamental matrix estimation (**F**) on kusvod2 (24 pairs), AdelaideRMF (19 pairs) and Multi-H (4 pairs) datasets, homography estimation (**H**) on homogr (16 pairs) and EVD (15 pairs) datasets, and essential matrix estimation (**E**) on the strecha dataset (467 pairs). In total, the testing included 545 image pairs. The datasets, the problem, the number of the image pairs (#) and the reported properties are shown in the first three columns. The other columns show the average results (100 runs on each image pair) of the competitor methods at 95% confidence. Columns with "$+\sigma$" show the results when the proposed σ-consensus was applied to the output of the method on its left. The mean geometric error (e_{avg}; in pixels) of the estimated model w.r.t. the manually selected inliers are written in each 1st row; the mean processing time (t, in milliseconds) and the required number of samples (s) are written in every 2nd and 3rd rows. In the 4th one, the proportion of failures, i.e. when the sough model is not found, is shown. The geometric error is the RMS Sampson distance for **F** and **E**, and the RMS re-projection error for **H** using the ground truth inlier set. The thresholds proposed in [11] were used. For MAGSAC, $\sigma_{max} = 10$ pixels.

6. Conclusion

A robust approach, called σ-consensus, was proposed for eliminating the need of a user-defined threshold by marginalizing over a range of noise scales. Also, due to not having a set of inliers, a new model quality function and termination criterion were proposed. Applying σ-consensus, we proposed two methods: first, MAGSAC applying σ-consensus to each of the models estimated from a minimal sample. The method is superior to the state-of-the-art in terms of geometric accuracy on publicly available real-world datasets for epipolar geometry (both **F** and **E**) and homography estimation. The method is often faster than other RANSAC variants in case of high outlier ratio. The proposed post-processing step applies σ-consensus only once: to polish the RANSAC output. The method nearly *always improved* the model quality on a wide range of vision problems without noticeable deterioration in processing time, i.e. at most a few milliseconds. We see no reason for not applying it after the robust estimation finished.

7. Acknowledgement

References

[1] O. Chum and J. Matas. Matching with PROSAC-progressive sample consensus. In *Computer Vision and Pattern Recognition*. IEEE, 2005. 1

[2] Ondřej Chum and Jiří Matas. Optimal randomized RANSAC. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1472–1482, 2008. 4, 5

[3] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. In *Joint Pattern Recognition Symposium*. Springer, 2003. 6

[4] O. Chum, T. Werner, and J. Matas. Epipolar geometry estimation via RANSAC benefits from the oriented epipolar constraint. In *International Conference on Pattern Recognition*, 2004. 6

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 1981. 1

[6] V. Fragoso, P. Sen, S. Rodriguez, and M. Turk. EVSAC: accelerating hypotheses generation by modeling matching scores with extreme value theory. In *International Conference on Computer Vision*, 2013. 1

[7] D. Ghosh and N. Kaabouch. A survey on image mosaicking techniques. *Journal of Visual Communication and Image Representation*, 2016. 1

[8] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 3, 6, 7

[9] R. I. Hartley. In defense of the eight-point algorithm. *Transactions on Pattern Analysis and Machine Intelligence*, 1997. 6

[10] H. Isack and Y. Boykov. Energy-based geometric multi-model fitting. *International Journal of Computer Vision*, 2012. 1

[11] K. Lebeda, J. Matas, and O. Chum. Fixing the locally optimized RANSAC. In *British Machine Vision Conference*. Citeseer, 2012. 6, 8

[12] D. G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer vision*. IEEE, 1999. 7

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 2004. 1

[14] D. Mishkin, J. Matas, and M. Perdoch. MODS: Fast and robust method for two-view matching. *Computer Vision and Image Understanding*, 2015. 1

[15] Lionel Moisan, Pierre Moulon, and Pascal Monasse. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line*, 2:56–73, 2012. 2, 5

[16] D. Nasuto and J. M. B. R. Craddock. NAPSAC: High noise, high dimensional robust estimation - its in the bag. 2002. 1

[17] T. T. Pham, T-J. Chin, K. Schindler, and D. Suter. Interacting geometric priors for robust multimodel fitting. *Transactions on Image Processing*, 2014. 1

[18] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *International Conference on Computer Vision*. IEEE, 1998. 1

[19] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J-M. Frahm. USAC: a universal framework for random sample consensus. *Transactions on Pattern Analysis and Machine Intelligence*, 2013. 4

[20] Martin Rais, Gabriele Facciolo, Enric Meinhardt-Llopis, Jean-Michel Morel, Antoni Buades, and Bartomeu Coll. Accurate motion estimation through random sample aggregated consensus. *CoRR*, abs/1701.05268, 2017. 5

[21] C. Sminchisescu, D. Metaxas, and S. Dickinson. Incremental model-based estimation using geometric constraints. *Pattern Analysis and Machine Intelligence*, 2005. 1

[22] Charles V. Stewart. Minpran: A new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):925–938, 1995. 2

[23] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: a probabilistic account. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 2004. 7

[24] P. H. S. Torr. Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision*, 50(1):35–61, 2002. 1

[25] P. H. S. Torr and D. W. Murray. Outlier detection and motion segmentation. In *Optical Tools for Manufacturing and Advanced Automation*. International Society for Optics and Photonics, 1993. 1

[26] P. H. S. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 2000. 1

[27] P. H. S. Torr, A. Zisserman, and S. J. Maybank. Robust detection of degenerate configurations while estimating the fundamental matrix. *Computer Vision and Image Understanding*, 1998. 1

[28] M. Zuliani, C. S. Kenney, and B. S. Manjunath. The multi-ransac algorithm and its application to detect planar homographies. In *International Conference on Image Processing*. IEEE, 2005. 1