Inverting RANSAC: Global Model Detection via Inlier Rate Estimation

Roee Litman* Simon Korman* Alex Bronstein Shai Avidan
School of Electrical Engineering, Tel-Aviv university

Abstract

This work presents a novel approach for detecting inliers in a given set of correspondences (matches). It does so without explicitly identifying any consensus set, based on a method for inlier rate estimation (IRE). Given such an estimator for the inlier rate, we also present an algorithm that detects a globally optimal transformation. We provide a theoretical analysis of the IRE method using a stochastic generative model on the continuous spaces of matches and transformations. This model allows rigorous investigation of the limits of our IRE method for the case of 2D-transformation, further giving bounds and insights for the more general case. Our theoretical analysis is validated empirically and is shown to hold in practice for the more general case of 2D-affinities. In addition, we show that the combined framework works on challenging cases of 2D-homography estimation, with very few and possibly noisy inliers, where RANSAC generally fails.

1. Introduction

The problem of image correspondence is a fundamental problem in computer vision, as it arises as a primitive in many tasks such as image retrieval, 3D reconstruction and panorama stitching. While some works solve these types of problems using direct methods [11, 5, 8], the vast majority of recent methods use large sets of matching (pairs of) points as their entry point, later discarding the content of the images. This is largely due to the tremendous improvement over the last decades in algorithms for detecting stable image feature points and representing them by descriptors that are designed for the task of matching [10, 13, 14].

The desired outcome of such a point matching process is that a large portion of the matches is accurate, while only a few of them (preferably none) can have arbitrarily bad errors. These two groups of matches are called inliers and outliers, respectively. The final step of image matching is therefore to robustly detect the “true” transformation underlying the inliers while ignoring the outliers. In practice, this is most commonly formulated as the consensus set maximization problem, where the goal is to find a maximal set of matches that agree on a model, up to some tolerance. Our method works in an opposite order: the inlier rate of matches is first estimated from the data and then, a model that minimizes the error of such a portion of inliers is searched for.

Figure 1. Our approach is “orthogonal” to RANSAC, which assumes a fixed error-threshold for inliers and then searches for a model that maximizes the inlier rate. Our method works in an opposite order: the inlier rate of matches is first estimated from the data and then, a model that minimizes the error of such a portion of inliers is searched for.

As a combinatorial metaphor of this phenomenon, consider a bag with $N$ balls, $k$ of which are white, and the remaining $N-k$ are black. In this metaphor, white and black balls correspond, respectively, to inliers and outliers. Also, a ‘selection’ of balls represents a transformation and $\nu$ is the

* equal contributors
number of ‘good’ possible selections. One has an estimate \( k \) of \( k \) and wishes to pick \( k \) balls with as many white ones as possible. If \( \hat{k} \) is an underestimate of \( k \), there are \( \binom{k}{\hat{k}} \) many options to do so. On the other hand, if \( \hat{k} \) is an overestimate of \( k \), all the \( k \) white balls must be selected along with \( k - \hat{k} \) additional black balls, for which there are \( \binom{N-k}{k-\hat{k}} \) options. These two cases coincide at \( k = \hat{k} \), where the number of options attains its minimum.

1.1. Prior work

A globally optimal solution for consensus set maximization can be obtained by naïvely going through all possible subsets of matches, a task of exponential magnitude. Nevertheless, many heuristics for its efficient approximation or full solution have been suggested in the literature, some with theoretical guarantees. These works, at large, can be divided into the following two categories.

RANSAC based techniques In RANSAC [4], the space of parameters is explored by repeatedly selecting random subsets of the matches for which a model hypothesis is fitted and then verified. A recent comprehensive survey and evaluation of RANSAC techniques by Raguram et al. [16] also suggests USAC – a uniform pipeline that combines several of the known extensions (e.g. [2, 12, 3]) in addition to many practical and computational considerations. USAC shows excellent results on a variety of transformation groups (e.g. Essential, Fundamental, Homography), in terms of accuracy, efficiency, and stability. Another interesting extension of RANSAC by Raguram et al. [17] claims to eliminate the need of the inlier-threshold input of RANSAC without harming exactness, at only a modest increase in runtime.

Global optimization techniques This line of works aims at overcoming the unpredictability of RANSAC based techniques, which is due to their inherent random nature. Similar to RANSAC, their formulation of consensus-set maximization uses a predefined inlier error threshold, which is a clear disadvantage. Ollson et al. [15] presented an approach based on theory from computational geometry. They give an \( O(k^{n+1}) \) polynomial time algorithm, for the case of \( k \) matches and transformation space of \( n \) DoF. This method could not be used in practice for spaces of more than a few DoF. Li et al. [9] proposed a solution that formulates the problem as a mixed integer program (MIP), which is generally NP-hard. However, they solve it exactly via relaxations, using a tailored branch-and-bound (BnB) scheme that involves solving a linear program at each node. While this approach generalizes nicely to other domains [1, 19, 20], unlike RANSAC it has not been shown to be efficient in challenging real-life cases where the portion of inliers is very small. The BnB scheme we propose involves much simpler calculations (computing and sorting match match errors) and can be applied successfully on such challenging cases.

1.2. Contributions

This paper has three main contributions. First, a scheme for efficiently sampling the space of transformations. Second, an algorithm for finding the best transformation for a set of matches, given the rate of inliers, with global guarantees. This algorithm has low practical applicability without our third, main, contribution - an algorithm (IRE) for estimating the rate of inliers in a given set of matches, without explicitly detecting them.

In addition, we present a rigorous analysis of the IRE algorithm and validate our analysis in several settings. We also show that our complete framework, which we term GMD, can work on challenging data with accuracy comparable to the state-of-the-art.

2. Method

Our algorithm gets as input a set of matches between a pair of images and a group of transformations to search through. As opposed to common practice, our philosophy is to first search for the rate of inliers (Section 2.3) and then search for the transformation with the lowest possible error over the specific rate of inliers (Section 2.2). These two components of the algorithm rely on a sampling regime of the space of transformations \( T \) (Section 2.1).

Preliminary definitions Let \( I_1 \) and \( I_2 \) be a pair of images, defined w.l.o.g. as 2D continuous entities on \([0, 1]^2\). A match \( m = (x_1, x_2) \) is an ordered pair of points \( x_1 \in I_1 \) and \( x_2 \in I_2 \) and thus we can denote the domain of matches to be the product of image domains, \( \mathcal{M} = [0, 1]^2 \times [0, 1]^2 \).

We denote by \( T \) a group of parametric transformations; we are mainly interested in the typical groups of transformations between pairs of images (i.e. functions from \( \mathbb{R}^2 \) to \( \mathbb{R}^2 \), with different degrees of freedom (DoF), such as Euclidean (3 DoF), Similarities (4 DoF), Affinities (6 DoF) and Homographies (8 DoF). In some of the cases we would like to further consider only a subspace of the group restricting, e.g., the maximum scale or the range of translation.

For any transformation \( t \in T \) and match \( m = (x_1, x_2) \in \mathcal{M} \) we define the error of the match \( m \) with respect to \( t \) to be the Euclidean distance in \( I_2 \):

\[
\text{err}(t, m) = \|x_2 - t(x_1)\|_2
\]  

(1)

Furthermore, we define a “worst-case” distance between any two transformations \( t_1, t_2 \in T \),

\[
d_T(t_1, t_2) = \max_{x_1 \in I_1} \|t_1(x_1) - t_2(x_1)\|_2.
\]  

(2)

This distance measures how far apart can any source image point be mapped by the two considered transformations.
The well-known Sampson error (see e.g. [6]) is obtained by replacing the max in (2) with an average. $d_T$ can easily be shown to be a metric, and will be used in the construction of a sampling of $T$.

2.1. Efficient sampling of $T$

In what follows, we construct a nearly uniform sample $S$ of the transformation group $T$. For a given resolution parameter $\varepsilon > 0$, we define a 2D Cartesian grid with step size $\sqrt{2}\varepsilon$ over the image $I_2$ (or a padded version of it). The sample $S_\varepsilon = \{t_1, \ldots, t_n\}$ is simply the subset of transformations in $T$ that map all four corners of $I_1$ to distinct grid points over $I_2$, as is illustrated in Figure 2. Under the definition of the distance $d_T$ (2), the covering and packing radii of $S_\varepsilon$ can be shown to be $\varepsilon$ and $\varepsilon/\sqrt{2}$, respectively, resulting in an $\varepsilon$-net. As is done in [8], the size of the $\varepsilon$-net can be shown to be $O(\varepsilon^{-n})$, for $T$ with $\eta$ DoF.

Another useful property of the sample $S_\varepsilon$ can be seen by examining the Voronoi tessellation it induces on the space $T$. For any match $m_i \in M$ and any $t_i \in S_\varepsilon$, the error $err(t_i, m_i)$ differs by at most $\varepsilon$ from the error $err(t, m)$ of any other transformation $t$ in the Voronoi cell of $t_i$. This follows from $d_T(t_i, t) < \varepsilon$ and the triangle inequality on $d_T$. Furthermore, this property holds for various statistics over the match errors in $M$, such as the mean, median or any other percentile.

2.2. Searching for an optimal transformation

The algorithm presented here finds an approximation $t_{\text{min}}$ for the optimal transformation $t^* \in T$, given an estimated inlier rate $\hat{p}$. By optimal we mean that the mean of the best $\hat{p}$-tile of match errors of $t^*$ is the lowest possible over all $t \in T$. To achieve this, we make use of the previously mentioned sample $S_\varepsilon$, and refine it only around promising regions in a branch-and-bound (BnB) manner. This method follows the lines of the template-matching method introduced in [8], and is summarized as Algorithm BnB.

Since the procedure can be repeated in a recursive manner, the error of the resulting $t_{\text{min}}$ can approach arbitrarily close to that of $t^*$. In the BnB process, we are guaranteed to never discard the Voronoi cell that contains $t^*$, centered at some $t \in S_\varepsilon$. This is due to the fact that the error $r_{\text{min}}(\hat{p})$ can not be lower than that of $t$ by more than $\varepsilon$. The complexity of one BnB iteration can be shown to be $O(k\varepsilon^{-n} + k\log k)$, for $k$ matches and $T$ with $\eta$ DoF.

2.3. Estimating the inlier rate

Since the inlier rate is seldom known in practice, we introduce Algorithm IRE, a practical procedure that finds an estimate $\hat{p}$ of the “true” inlier rate $p^*$ in the set of matches $M$. In order to do so, we introduce a quantity (denoted by $v_\varepsilon(p)$ in the algorithm) that depends on the sample density $\varepsilon$ and is a function of the inlier rate $p$. This quantity is very easy to compute and the main insight of the paper, is that it attains a minimum at the “true” inlier rate $p^*$. In Section 3, we give an extensive theoretical analysis of the existence of

**Algorithm 1:** Finding the best transformation through branch-and-bound (BnB).

1. Construct a sample $S_\varepsilon$ of $T$ (Section 2.1).
2. Compute an error matrix $E$, with entries $e_{ij} = err(m_i, t_j)$ for each $m_i \in M$ (rows) and each $t_j \in S_\varepsilon$ (columns).
3. Sort each column of $E$ by increasing error (as a result every row represents a percentile $\hat{p}$).
4. Replace each column by the cumulative average of its entries (as a result each entry holds the average of match errors from lower percentiles).
5. Extract from $E$ the row $e_{\hat{p}}$, that corresponds to the percentile $\hat{p}$.
6. Find the transformation $t_{\text{min}}$, that attains the minimal error in $e_{\hat{p}}$, denoted as $r_{\text{min}}(\hat{p})$.

branch-and-bound extension

7. If $\varepsilon$ is low enough: terminate and return $t_{\text{min}}$.
8. Discard all transformations (columns) that have $e_{\hat{p}} > r_{\text{min}}(\hat{p}) + \varepsilon$.
9. Replace the remaining samples of $S_\varepsilon$ with their children in $S_{\varepsilon/2}$.
10. Set $\varepsilon \leftarrow \varepsilon/2$ and go to step 2.
is drawn from the conditional distribution of the outliers, both currently left unspecified.

3.2. Probabilistic interpretation of the model

Having defined a distribution of point matches, we can now measure probabilities over match errors with respect to some transformation \( t \in \mathcal{T} \). Specifically, we are interested in the probability of a match \( m \) to have an error below some threshold \( r \),

\[
p_t(r) = \Pr\{\text{err}(t, m) \leq r\}
\]

where the probability is taken over the distribution \( f_m \) of matches \( m \). Using this notation, it is now clear that the distribution \( f_m \) (specifically \( f_2 \)) was defined so that \( p_t(r^*) = p^* \). The probability \( p_t(r) \) can be computed by marginalizing over \( x_1 \in I_1 \),

\[
p_t(r) = \int_{x_1 \in I_1} f_1(x_1)q_t(r|x_1) \, dx_1
\]

where \( q_t(r|x_1) \) is the conditional probability for a match with a specific source point \( x_1 \) to have an error less than \( r \). Substituting the distribution \( f_2 \) defined in (4), yields

\[
q_t(r|x_1) = \int_{\mathcal{B}_r(t(x_1))} f_2(x_2|x_1) \, dx_2
\]

where the integration domain is the Euclidean ball of radius \( r \) centered at the target point \( t(x_1) \) in the image \( I_2 \).

It is worth while pointing out the fact that there is a monotonic non-decreasing relation between the error radius \( r \) and the probability \( p_t(r) \): the higher the error threshold is the higher the probability of a match error to be within the threshold. This relation enables us to introduce an equivalent term for the error radius \( r \) for which \( p_t(r) = p \).

\[
r_t(p) = \min r \quad \text{s.t.} \quad p_t(r) = p.
\]

3.3. Inlier rate estimation

In this section, we formulate the probabilistic version of Algorithm IRE. This includes, the definition of the continuous counterparts of the vectors \( v_e \) and \( r_{\text{min}} \). We first define \( r_{\text{min}}(p) \) to be the best attainable error for any transformation \( t \in \mathcal{T} \) that “captures” matches with probability \( p \),
It is easy to see that \( r_{\text{min}}(p^*) = r^* \), and that this value is achieved, possibly among others, by \( t^* \) (otherwise the inliers are governed by some other, more prominent, transformation). To define the continuous counterpart of \( \nu_* \), we can no longer use the sampling resolution \( \varepsilon \). Instead, we define an error tolerance \( \delta \). In Algorithm \textsc{Iter} implicitly link the two parameters by setting \( \delta = \varepsilon \), which is used throughout our experiments; however, we stress that the exact relation between these parameters requires additional research. For every \( p \) and \( \delta > 0 \), we define

\[
\Omega_\delta(p) = \{ t \in \mathcal{T} : r_t(p) \leq r_{\text{min}}(p) + \delta \},
\]

as the subset of \( \mathcal{T} \) of transformations \( t \) with error radius \( r_t(p) \) that is at most \( \delta \) larger than the optimal radius \( r_{\text{min}}(p) \). We take the normalized volume of the subset \( \Omega_\delta(p) \) to be our indicative quantity for estimating \( p^* \):

\[
V_\delta(p) = \frac{\text{Vol}(\Omega_\delta(p))}{\text{Vol}(\mathcal{T})}.
\]

We can now formulate our main claim regarding the behavior of our measure \( V_\delta(p) \) around the true inlier rate \( p^* \):

**Proposition 1.** If both \( f_{\text{in}} \) and \( f_{\text{out}} \) are uniform distributions and \( \mathcal{T} \) is the space of 2D translations, then

\[
p^* = \arg \min_p V_\delta(p).
\]

**Proof.** Let us begin by spelling out the assumption of uniform \( f_{\text{in}} \) and \( f_{\text{out}} \). For \( f_{\text{in}} \) we assume that the target points \( x_2 \) of inlier matches are distributed uniformly in \( B_{r^*}(t^*(x_1)) \), i.e., on a ball of radius \( r^* \) around \( t^*(x_1) \). For \( f_{\text{out}} \) we assume that the target points \( x_2 \) are distributed uniformly on the entire image, except \( B_{r^*}(t^*(x_1)) \). We denote these two constant probability densities as \( \rho_{\text{in}} = (\pi r^{*2})^{-1} \) and \( \rho_{\text{out}} = (1 - \pi r^{*2})^{-1} \), respectively.

The main advantage of assuming uniform distributions is that probability calculations reduce to area calculations. Specifically, looking at Figure 3, the probability \( q_t(r_{\text{min}}(p) + \delta | x_1) \) is the one captured in the red ball \( B_{r_{\text{min}}(p) + \delta}(t(x_1)) \). Its calculation can be broken down to the inlier area (intersection of red and blue balls) weighted by \( \rho_{\text{in}} \) and the outlier area (rest of the red ball) weighted by \( \rho_{\text{out}} \). It follows that \( q_t(r_{\text{min}}(p) + \delta | x_1) \) depends only on the distance \( d_{t} = \| t(x_1) - t^*(x_1) \|_2 \) between the ball centers in \( I_2 \), marked in green in Figure 3. Assuming that \( \rho_{\text{in}} \cdot p^* > \rho_{\text{out}} \cdot (1 - p^*) \) (prominence of the inlier model), the probability \( q_t(r_{\text{min}}(p) + \delta | x_1) \) decreases as \( d_{t} \) grows.

An equivalent way of looking at \( \Omega_\delta(p) \) follows from the definition of \( r_t(p) \) in (8)

\[
\Omega_\delta(p) = \{ t \in \mathcal{T} : p_t(r_{\text{min}}(p) + \delta) \geq p \},
\]

which leads to a sufficient (but not necessary) condition for a certain transformation \( t \) to be in \( \Omega_\delta(p) \):

\[
q_t(r_{\text{min}}(p) + \delta | x_1) \geq p, \; \forall x_1 \in I_1. \tag{13}
\]

In other words, a ball of radius \( r_{\text{min}}(p) + \delta \) centered around \( t(x_1) \) should contain a probability of at least \( p \), for all \( x_1 \). In the case of 2D translations, the latter condition is also necessary. To show that, we observe that \( d_t \) is constant over all \( x_1 \in I_1 \), and so is the probability \( q_t(r_{\text{min}}(p) + \delta | x_1) \) that depends on it. The expression for \( p_t(r) \) in (6) yields \( p_t(r) = q_t(r | x_1) \) for every \( x_1 \) regardless of \( f_1 \), and condition (13) holds iff \( t \in \Omega_\delta(p) \). Since \( d_t \) is constant over all \( x_1 \in I_1 \), \( \Omega_\delta(p) \) can be defined as

\[
\Omega_\delta(p) = \{ t : d_T(t^*, t) \leq d_{\text{max}}(p) \}, \tag{14}
\]

using the distance \( d_T \) from (2), where \( d_{\text{max}}(p) \) denotes the maximal distance \( d_T \) at which inequality (13) still holds.
With this interpretation of $\Omega_S(p)$ as a ball of radius $d_{\max}(p)$ (in transformation space), its volume increases monotonically with the radius, and therefore $V_S(p)$ attains a minimum at $p^*$ iff $d_{\max}(p)$ does so.

To show the existence of the latter minimum, we study the behavior of $d_{\max}(p)$ around $p^*$. The probability $q_t(r_{\min}(p) + \delta \mid x_1)$ from (13) can be expressed as

$$q_t(r_{\min}(p) + \delta \mid x_1) = \ldots$$
$$\rho_{\min} \cdot p^* \cdot \psi(r^*, r_{\min}(p) + \delta, d_{\max}(p)) + \ldots$$
$$\rho_{\text{out}} \cdot (1 - p^*) \cdot [\pi (r_{\min}(p) + \delta)^2 - \psi(r^*, r_{\min}(p) + \delta, d_{\max}(p))]$$

where $\psi(r_1, r_2, d)$ is the intersection area of circles of radius $r_1$ and $r_2$ with centers at distance $d$ (see Supplementary Materials for closed-form expressions of $r_{\min}$ and $\psi$). The value of $d_{\max}(p)$ is found by plugging inequality (13) into (15). By performing implicit differentiation on the result, we show that $d_{\max}(p)$ attains a local minimum at $p = p^*$ (full derivations in Supplementary Materials), for a large range of values of $p^*$ and $r^*$, illustrated in Figure 4.

The simple form of $\Omega_S(p)$ obtained in (14) for 2D translation leads to a closed-form expression for the normalized volume, $V_S(p) = \pi d_{\max}(p)^2 / \text{Vol}(T)$, following (10). As expected for a 2D space, the volume grows quadratically with the radius. Figure 5 shows the (analytically computed) quantities $r_{\min}(p)$, $d_{\max}(p)$ and $V_S(p)$ for 2D translations as a function of $p$, for a specific setting of $p^*$, $r^*$ and $\delta$. As expected, both $d_{\max}(p)$ and $V_S(p)$ attain a minimum at $p^*$.

While Proposition 1 is limited to only one type of distribution (uniform) and one type of transformations (2D translation), we conjecture that it holds for a much wider range of settings and provide evidence for this in Sections 4.2 and 4.3. We discuss in Section 5 possible extensions to some more complex real-life cases and their challenges.

4. Experimental Results

We present a series of three experiments that examine the proposed method in a gradual manner, going from theoretic to real-life cases.

4.1. 2D translation - Synthetic data

We first wish to validate the formula for $V_S(p)$ for the case of 2D translations, presented right after the proof of Proposition 1. Since the space of 2D translations has only two DoF we can sample it densely, and obtain a close-to-continuous approximation of the size $V_S(p)$. Figure 6 shows results on a single instance of the problem generated according to our model, with inlier rate $p^* = 8\%$ and inlier noise $\epsilon^*$ of $5\%$ of the size of $I_2$. We compare the volume $V_S(p)$ as obtained from empirical measurements (solid lines) to the theory (dashed lines) for several values of $\delta$ (color coded). There is an evident match between theory and practice at a certain interval around the true inlier rate $p^*$ (black dashed line). The extent of this interval diminishes with the increase in $\delta$, a phenomenon we discuss in the Supplementary Materials. In addition, note that all the curves attain a minimum at $p^*$, as predicted by Proposition 1, even for high $\delta$ values, in accordance with the solution regions in Figure 4.

4.2. 2D affine - Synthetic data

While our theoretic analysis was developed for a continuous space of transformations, in practice both Algorithm BnB and Algorithm IRE rely on discrete samplings $S_\epsilon$ of the space $T$. The sampling density depends mainly on memory and time considerations, and tends to be effectively coarser for $T$ with many DoF. In this experiment, we examine how our IRE method works under deteriorating sampling resolutions on the 2D-Affine group (6 DoF). Coarse sampling causes the method to deviate from the continuous version in two ways. First, the calculation of $r_{\min}$ is an approximation of $r_{\min}(p)$; however, since $S_\epsilon$ is an $\varepsilon$-covering we incur an additive error of at most $\varepsilon$. Second, $V_\delta(p) / |S_\epsilon|$ approximates the normalized volume $V_S(\varepsilon)$. The fact that the sample is relatively uniform in the distance $d_t(S_\epsilon$ has similar covering and packing radii), ensures that sample counting approximates well the volume.

Figure 7 presents results on two instances of the experiment, which were generated in a way similar to the previ-
Iₙ, I₂, Iₙ, Iₙ, Iₙ, Iₙ.

Figure 7. Sensitivity to sampling of a synthetic 2D affine group example. We show here results for two challenging examples. **Left example** (low inlier rate p⁺): p⁺ = 1% and r⁺ = 1%. **Right example** (high inlier noise r⁺): p⁺ = 16% and r⁺ = 8%. In each example: The image I₁ (top left) is mapped into image I₂ by an affine transformation (solid black parallelogram). 500 matches were generated according to our model with the mentioned p⁺ and r⁺ (inlier matches are in blue). We calculate vₚ using a sequence of step-sizes εᵢ (color coded plots), and it is evident that the location of the minimum (black cross) stays roughly around p⁺.

The first example is extreme in terms of the inlier-rate and the other is extreme in terms of the inlier-noise. Each curve (color coded by sample density εᵢ) shows our approximation vₚ(εᵢ) of the normalized volume Vₚ(εᵢ). In both cases, at a range of sampling resolutions, the minimum value is obtained at the true inlier rate or relatively close to it. This stability of the IRE algorithm under changes of the error tolerance δ is also indicated in Figures 4 (right) and 6.

### 4.3. 2D homography - Real data

In this experiment, we test our algorithm on two datasets containing very challenging image pairs, as the inliers are noisy and their rate can fall well below 10%. We also compare our results to USAC [16], a state-of-the-art RANSAC method, with a publicly available implementation.

#### Datasets

The first one was presented by Mikolajczyk et al. [14], and was originally constructed to benchmark feature detectors and descriptors. Here we use 5 of the sequences - each containing 6 images where a planar object undergoes a gradually increasing view point change. As was suggested in the dataset, we use the pairs 1-2, 1-3, 1-4, 1-5, 1-6, for which a ground-truth homography is provided. The second dataset was used by Raguram et al. [16] to benchmark the USAC algorithm against other RANSAC methods. We use a portion of this dataset which includes image pairs related by view-point change, described by a homography. Since the USAC algorithm requires an inlier error-threshold, we ran it for each integer threshold from 2 pixels (the recommended default) up to 30 pixels and took the run of the lowest threshold for which the run succeeded. For image pairs in [16] there was no ground truth provided, and we created one manually. In both datasets, the ground-truth is accurate up to 1 pixel, which is sufficient for comparison.

#### Implementation details

For all images, we generated correspondences based on matching SIFT descriptors, using the VLFeat library [18]. We used an initial sampling resolution ε that equals a third of the minimal image dimension. Our Matlab implementation of Algorithm BnB typically takes less than 10 seconds for an image pair on a modest PC. The runtime of Algorithm IRE is negligible as it reuses most of the calculations done in the former.

In this experiment, we applied two heuristics after step 6 of Algorithm BnB in order to accelerate it (without compromising the guarantees). First, we performed a depth-first-search around tₘᵢₙ, possibly providing a lower rₘᵢₙ(δ) which allows us to discard more samples. This heuristic is somewhat similar to the one used in [9]. Second, we perform a local optimization of tₘᵢₙ by reweighted least squares on the inliers. This heuristic may also lower rₘᵢₙ(δ), and is somewhat similar to the LO-RANSAC [3] extension used in USAC. In addition, the local optimization improves accuracy. The result is closer to the ground-truth transformation (i.e. accurate in terms of Sampson error [6]).

#### Results

Figure 8 shows results on three image pairs from [16]. As can be seen in the middle column, the minimum of vₚ(δ) is prominent in all cases, and its location is close to the “saturation” level of the (green) CDF of match errors w.r.t. the ground-truth (shown on the right) - an indication of the correctness of the detection. In addition, results for USAC are shown by the red circles on the right for thresholds in the interval [2,30] at steps of 3. Notice that in the second row there are very few inliers with error < 2 and
Table 1. Sampson error of homography estimation in five viewpoints of four scenes from [14] (each row corresponds to a scene and each column to a pair of images, with the strength of the viewpoint transformation increasing from left to right). Errors are reported w.r.t. ground truth, which may be up to one pixel inaccurate.

Table 2. Homography estimation inlier rates and match errors on the graffiti and graffiti-5 sequences from [14]. For each method (GMD and USAC), the detected inlier rate is shown in the first row, and the median±std of the inlier match errors is in the second. USAC achieves lower error levels since it typically classifies less matches as inliers. Notice that GMD succeeded in two cases where USAC failed.

5. Discussion and Future Work

We have presented and new approach to detecting a model from matches, where the rate of inliers is estimated first and only then the best transformation is searched for. The method seems to perform very well in practice on challenging cases of homography estimation, even if the theoretical background is currently limited. We made two restrictive assumptions in Proposition 1, and while alleviating these assumptions is deferred to future work, in what follows we briefly discuss the main implications.
References


