GMD: Global Model Detection via Inlier Rate Estimation

Roee Litman, Simon Korman, Alex Bronstein, Shai Avidan
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-translation, 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.

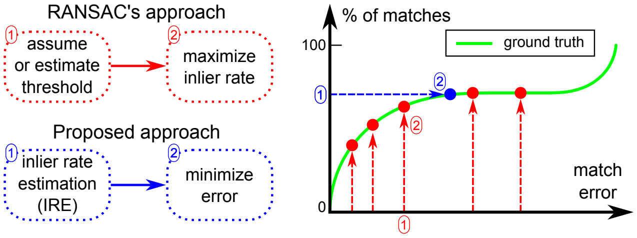

Figure 1. **Our approach is "orthogonal" to RANSAC,** which assumes a fixed error-threshold for inliers and then searches for a model that *maximizes the inlier rate* Our method works in an opposite order: the inlier rate of matches is first estimated from the data and then, a model that *minimizes the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

This work presents a different approach, as is illustrated in Figure 1. The green curve is a cumulative error distribution of matches between a pair of images, under a ground-truth transformation. In RANSAC (and any other consensus-set-maximization approach), a fixed error threshold is chosen and a model with a maximal number of inliers within the threshold is searched for (depicted by vertical red lines, for different thresholds). Our approach, on the other hand, first estimates the true inlier rate of the matches (about 70$\%$ in this example) and then searches for a model with lowest possible match errors over the detected portion of matches (depicted by the blue horizontal arrow).

The portion of inliers and the noise level of inlier matches are generally unknown, and any a-priori choice of error threshold is rather arbitrary. Our inlier rate estimation (IRE) method makes a principled prediction based on minimizing an indicative quantity, denoted $\mathbf{v}(p)$, over any possible inlier rate p. $\mathbf{v}$ 'counts' the number of transformations (or portion of transformation space) that have a p-tile error 'similar' to the best one possible. It turns out that $\mathbf{v}$ has a very particular behavior around the true inlier rate, where it attains a surprisingly clear minimum.

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 $\mathbf{v}$ is the number of 'good' possible selections. One has an estimate $\hat{k}$ of k and wishes to pick $\hat{k}$ 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 $\hat{k} - k$ additional black balls, for which there are $\binom{N-k}{\hat{k}-k}$ 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 $O(k^{\eta+1})$ polynomial time algorithm, for the case of k matches and transformation space of η 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 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 $\mathcal{T}$ (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* $m = (\mathbf{x}_1, \mathbf{x}_2)$ is an ordered pair of points $\mathbf{x}_1 \in I_1$ and $\mathbf{x}_2 \in I_2$ 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 $\mathcal{T}$ 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 $t \in \mathcal{T}$ and match $m = (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{M}$ we define the *error* of the match m with respect to t to be the Euclidean distance in I_2:

$$\mathrm{err}(t, m) = \|\mathbf{x}_2 - t(\mathbf{x}_1)\|_2 \tag{1}$$

Furthermore, we define a "worst-case" distance between any two transformations $t_1, t_2 \in \mathcal{T}$,

$$d_{\mathcal{T}}(t_1, t_2) = \max_{\mathbf{x}_1 \in I_1} \|t_1(\mathbf{x}_1) - t_2(\mathbf{x}_1)\|_2. \tag{2}$$

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. $d_{\mathcal{T}}$ can easily be shown to be a metric, and will be used in the construction of a sampling of $\mathcal{T}$.

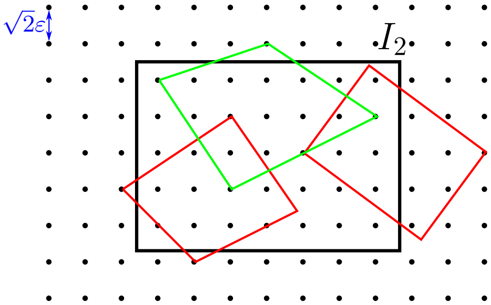

Figure 2. **Construction of the sampling $\mathcal{S}_\varepsilon$:** A Cartesian grid (black points) with step size $\sqrt{2}\varepsilon$ over the image I_2 (black rectangle) defines the sample. For a transformation to be a part of $\mathcal{S}_\varepsilon$, it has to map *all* four corners of I_1 to four points on the grid. An example of a valid map is shown in green, and two non-valid maps are shown in red.

2.1. Efficient sampling of $\mathcal{T}$

In what follows, we construct a nearly uniform sample $\mathcal{S}$ of the transformation group $\mathcal{T}$. 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 $\mathcal{S}_\varepsilon = \{t_1, \ldots, t_n\}$ is simply the subset of transformations in $\mathcal{T}$ that map all four corners of I_1 to distinct grid points[1] over I_2, as is illustrated in Figure 2. Under the definition of the distance $d_\mathcal{T}$ (2), the covering and packing radii of S_ε can be shown to be ε and $\varepsilon/\sqrt{2}$, respectively[2], resulting in an ε*-net*. As is done in [8], the size of the ε-net can be shown to be $O(\varepsilon^{-\eta})$, for $\mathcal{T}$ with η DoF.

Another useful property of the sample $\mathcal{S}_\varepsilon$ can be seen by examining the Voronoi tessellation it induces on the space $\mathcal{T}$. For any match $m_j \in M$ and any $t_i \in \mathcal{S}_\varepsilon$, the error $\mathrm{err}(t_i, m_j)$ differs by at most ε from the error $\mathrm{err}(t, m_j)$ of any other transformation t in the Voronoi cell of t_i. This follows from $d_\mathcal{T}(t_i, t) < \varepsilon$ and the triangle inequality on $d_\mathcal{T}$. 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 $t_{\min}$ for the *optimal* transformation $t^\star \in \mathcal{T}$, given an estimated inlier rate $\hat{p}$. By optimal we mean that the mean of the best $\hat{p}$-tile of match errors of $t^\star$ is the lowest possible over all $t \in \mathcal{T}$. To achieve this, we make use of the previously mentioned sample $\mathcal{S}_\varepsilon$, 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.

```
input: Inlier-rate estimate $\hat{p}$; matches $M$; resolution $\varepsilon$;
output: Estimate $t_{\min}$ of the best transformation $t^\star$

1. Construct a sample $\mathcal{S}_\varepsilon$ of $\mathcal{T}$ (Section 2.1).
2. Compute an error matrix $\mathbf{E}$, with entries
   $e_{ij} = \mathrm{err}(t_i, m_j)$ for each $t_i \in \mathcal{S}_\varepsilon$ (rows) and each
   $m_j \in M$ (columns).
3. Sort each column of $\mathbf{E}$ by increasing error
   (as a result every row represents a percentile $p$).
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 $\mathbf{E}$ the row $\mathbf{e}_{\hat{p}}$, that corresponds to the
   percentile $\hat{p}$.
6. Find the transformation $t_{\min}$, that attains the minimal error in $\mathbf{e}_{\hat{p}}$, denoted as $r_{\min}(\hat{p})$.
branch-and-bound extension
7. If $\varepsilon$ is low enough: terminate and return $t_{\min}$.
8. Discard all transformations (rows) that have
   $\mathbf{e}_{\hat{p}} > r_{\min}(\hat{p}) + \varepsilon$.
9. Replace the remaining samples of $\mathcal{S}_\varepsilon$ with their
   children in $\mathcal{S}_{\varepsilon/2}$.
10. Set $\varepsilon \leftarrow \varepsilon/2$ and go to step 2.
```

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

Since the procedure can be repeated in a recursive manner, the error of the resulting $t_{\min}$ can approach arbitrarily close to that of $t^\star$. In the BnB process, we are guaranteed to never discard the Voronoi cell that contains $t^\star$, centered at some $t \in \mathcal{S}_\varepsilon$. This is due to the fact that the error $r_{\min}(\hat{p})$ can not be lower than that of t by more than ε. The complexity of one BnB iteration can be shown to be $O(k\varepsilon^{-\eta} + k\log k)$, for k matches and $\mathcal{T}$ with η 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 $p^\star$ in the set of matches M. In order to do so, we introduce a quantity (denoted by $\mathbf{v}_\varepsilon(p)$ in the algorithm) that depends on the sample density ε 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 $p^\star$. In Section 3, we give an extensive theoretical analysis of the existence of this minimum.

To understand the idea behind the method, for any inlier rate p, think of the transformation $t_{\min}(p)$ - the one which attains the minimal error over any $p\%$ of the matches. It turns out that the closer p is to the true rate $p^\star$, the fewer

[1] This requirement prevents usage for spaces that allow only "rigid" rotations like similarity transforms, but it is easy to define an alternative sampling for these cases, as is done in [7].

[2] This means that any two samples in $\mathcal{S}_\varepsilon$ are at least $\sqrt{2}\varepsilon$ apart and that any $t \in \mathcal{T}$ has at least one sample in $\mathcal{S}$ that is at most ε away from it.

```
input: Set of matches $M$; Sample resolution $\varepsilon$
output: Estimate $\hat{p}$ of the "true" inlier rate $p^\star$

1. Construct a sorted error matrix $\mathbf{E}$
   (steps 1-3 in Algorithm BnB).
2. Calculate a column vector $\mathbf{r}_{\min}$ by taking the minimal error from each row of $\mathbf{E}$.
3. Calculate a column vector $\mathbf{v}_\varepsilon(p)$ by counting the number of entries in each row of $\mathbf{E}$ that are at most $\varepsilon$ more than their respective value in $\mathbf{r}_{\min}$.
4. Take $\hat{p}$ to be the relative location of the minimal value in $\mathbf{v}_\varepsilon(p)$.
```

Algorithm 2: Inlier Rate Estimation (IRE).

transformations there are with error "similar" to that of $t_{\min}(p)$. Our measure $\mathbf{v}_\varepsilon(p)$ counts the number of transformations that can explain $p\%$ of the matches with an error that is within a tolerance of ε from that of $t_{\min}(p)$.

3. IRE Theoretic Justification

In this section we give the theoretical background behind our IRE algorithm from Section 2.3. Our formulation of the problem uses a generative model in which the set of matches is drawn from a distribution, which generates both inliers and outliers. The probabilistic properties of this model will allow us to obtain bounds on the quantity $\mathbf{v}_\varepsilon$ and to assert that it has a local minimum around the "true" inlier rate $p^\star$.

3.1. Generative model for matches

Our formulation of the distribution of matches f_m below, is governed by 3 main factors. First, the *inlier rate* $p^\star$, which is the probability of a match to be an inlier rather than an outlier. Second, the "*true*" transformation $t^\star \in \mathcal{T}$ which is used to generate inlier matches. Finally, a maximal noise magnitude $r^\star$ that may be added to inlier locations.

A pair of matching points $m = (\mathbf{x}_1, \mathbf{x}_2) \in \mathcal{M}$ is drawn from the distribution

$$f_m(\mathbf{x}_1, \mathbf{x}_2) = f_1(\mathbf{x}_1) f_2(\mathbf{x}_2|\mathbf{x}_1), \tag{3}$$

where first the point $\mathbf{x}_1$ is drawn according to some arbitrary distribution[3] $f_1(\mathbf{x}_1)$ on I_1, followed by the point $\mathbf{x}_2$ which is drawn from the conditional distribution

$$f_2(\mathbf{x}_2|\mathbf{x}_1) = p^\star \cdot f_{\text{in}}(\mathbf{x}_2|\mathbf{x}_1) + (1-p^\star) \cdot f_{\text{out}}(\mathbf{x}_2|\mathbf{x}_1)\,. \tag{4}$$

In our model, the inlier distribution $f_{\text{in}}(\mathbf{x}_2|\mathbf{x}_1)$ places $\mathbf{x}_2$ at the location $t^\star(\mathbf{x}_1)$ in I_2 and adds to it random noise with maximal magnitude of $r^\star$, such that $\text{err}(t^\star, m) \le r^\star$. The outlier distribution $f_{\text{out}}(\mathbf{x}_2|\mathbf{x}_1)$, places the point $\mathbf{x}_2$ at random in I_2. We assume w.l.o.g. that $\mathbf{x}_2$ is placed at distance of at least $r^\star$ from $t^\star(\mathbf{x}_1)$ such that $\text{err}(t^\star, m) > r^\star$, since otherwise such a match can be considered to be an inlier.

As is done in different formulations regarding distributions of inliers and outliers (e.g. in consensus set maximization problems), we will later make assumptions on the kind of noise that is added to the inliers and on the specific 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,

$$p_t(r) = \text{P}\{\text{err}(t, m) \le r\}\ , \tag{5}$$

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 $p_{t^\star}(r^\star) = p^\star$. The probability $p_t(r)$ can be computed by marginalizing over $\mathbf{x}_1 \in I_1$,

$$p_t(r) = \int_{\mathbf{x}_1 \in I_1} f_1(\mathbf{x}_1) q_t(r|\mathbf{x}_1)\, d\mathbf{x}_1 \tag{6}$$

where $q_t(r|\mathbf{x}_1)$ is the conditional probability for a match with a specific source point $\mathbf{x}_1$ to have an error less than r. Substituting the distribution f_2 defined in (4), yields

$$q_t(r|\mathbf{x}_1) = \int_{\text{B}_r(t(\mathbf{x}_1))} f_2(\mathbf{x}_2|\mathbf{x}_1)\, d\mathbf{x}_2 \tag{7}$$

where the integration domain is the Euclidean ball of radius r centered at the target point $t(\mathbf{x}_1)$ 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 $p_t(r) = p$,

$$r_t(p) = \min\ r \quad \text{s.t.} \quad p_t(r) = p\ . \tag{8}$$

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 $\mathbf{v}_\varepsilon$ and $\mathbf{r}_{\min}$. We first define $r_{\min}(p)$ to be the best attainable error for any transformation $t \in \mathcal{T}$ that "captures" matches with probability p,

$$r_{\min}(p) = \min_{t \in \mathcal{T}} r_t(p).$$

[3] For example the distribution of occurrences of interest points in an image. We ignore the case of inlier points that are mapped under $t^\star$ outside of I_2, for which the inlier distribution should be zero.

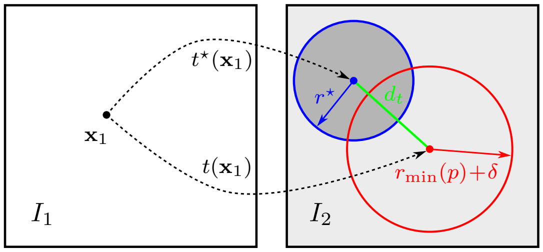

Figure 3. **Characterizing the transformation t by the distance d_t from $t^\star$**: For a specific point $\mathbf{x}_1 \in I_1$, we look at its possible target locations in I_2. The blue ball $\mathrm{B}_{r^\star}(t^\star(\mathbf{x}_1))$ has probability $p^\star$ since it contains the inlier distribution entirely. Any transformation t is associated with a red ball $\mathrm{B}_{r_{\min}(p)+\delta}(t(\mathbf{x}_1))$. For t to be in $\Omega_\delta(p)$ the red ball should contain a probability of at least p.

It is easy to see that $r_{\min}(p^\star) = r^\star$, and that this value is achieved, possibly among others, by $t^\star$ (otherwise the inliers are governed by some other, more prominent, transformation). To define the continuous counterpart of $\mathbf{v}_\varepsilon$, we can no longer use the sampling resolution ε. Instead, we define an error tolerance δ. In Algorithm IRE we 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

$$\Omega_\delta(p) = \{t \in \mathcal{T} : r_t(p) \le r_{\min}(p) + \delta\}, \qquad (9)$$

as the subset of $\mathcal{T}$ of transformations t with error radius $r_t(p)$ that is at most δ larger than the optimal radius $r_{\min}(p)$. We take the normalized *volume*[4] of the subset $\Omega_\delta(p)$ to be our indicative quantity for estimating $p^\star$:

$$V_\delta(p) = \mathrm{Vol}(\Omega_\delta(p)) \,/\, \mathrm{Vol}(\mathcal{T}). \qquad (10)$$

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

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

$$p^\star = \operatorname*{argmin}_{p} V_\delta(p) \;. \qquad (11)$$

Proof. Let us begin by spelling out the assumption of uniform f_{in} and f_{out}. For f_{in} we assume that the target points $\mathbf{x}_2$ of inlier matches are distributed uniformly in $\mathrm{B}_{r^\star}(t^\star(\mathbf{x}_1))$, i.e., on a ball of radius $r^\star$ around $t^\star(\mathbf{x}_1)$. For f_{out} we assume that the target points $\mathbf{x}_2$ are distributed uniformly on the entire image, except $\mathrm{B}_{r^\star}(t^\star(\mathbf{x}_1))$. We denote these two constant probability densities as $\rho_{\text{in}} = (\pi {r^\star}^2)^{-1}$ and $\rho_{\text{out}} = (1 - \pi {r^\star}^2)^{-1}$, respectively.

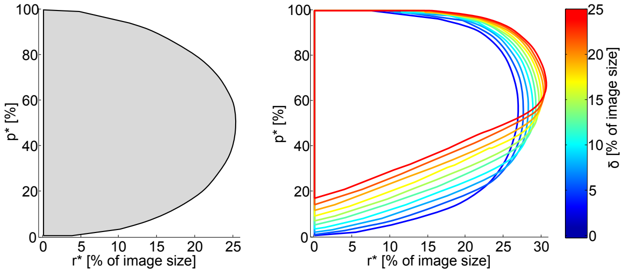

Figure 4. **Regions in the $(r^\star, p^\star)$ plane where $d_{\max}(p)$ attains a minimum at $p^\star$. Left:** The region for $\delta \ll 1$, the interior of the region is colored in gray. Note that the minimum exists for a large range of values of $p^\star$ and $r^\star$. **Right:** Regions for large δ values (interior not colored). The fact the the region is large even for large values of δ give intuition on the performance with coarse sampling of $\mathcal{T}$. See Supplementary Materials for a detailed discussion.

The main advantage of assuming uniform distributions is that probability calculations are reduced to area calculations. Specifically, looking at Figure 3, the probability $q_t(r_{\min}(p) + \delta \,|\, \mathbf{x}_1)$ is the one captured in the red ball $\mathrm{B}_{r_{\min}(p)+\delta}(t(\mathbf{x}_1))$. The calculation can be broken down to the inlier area (intersection between the red and blue balls) weighted by ρ_{in} and the outlier area (the rest of the red ball) weighted by ρ_{out}. It follows that $q_t(r_{\min}(p){+}\delta \,|\, \mathbf{x}_1)$ depends only on the distance $d_t = \|t(\mathbf{x}_1) - t^\star(\mathbf{x}_1)\|_2$ between the ball centers in I_2, marked in green in Figure 3. Under the necessary assumption that $\rho_{\text{in}} \cdot p^\star > \rho_{\text{out}} \cdot (1 - p^\star)$, the probability $q_t(r_{\min}(p){+}\delta \,|\, \mathbf{x}_1)$ decreases as d_t grows.

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

$$\Omega_\delta(p) = \{t \in \mathcal{T} : p_t(r_{\min}(p){+}\delta) \ge p\}\;, \qquad (12)$$

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

$$q_t(r_{\min}(p){+}\delta \,|\, \mathbf{x}_1) \ge p, \;\forall \mathbf{x}_1 \in I_1\,. \qquad (13)$$

In other words, a ball of radius $r_{\min}(p){+}\delta$ centered around $t(\mathbf{x}_1)$ should contain a probability of at least p, for all $\mathbf{x}_1$. In the case of 2D translations, the latter condition is also necessary. To show that, we observe that d_t is constant over all $\mathbf{x}_1 \in I_1$, and so is the probability $q_t(r_{\min}(p){+}\delta \,|\, \mathbf{x}_1)$ that depends on it. The expression for $p_t(r)$ in (6) yields $p_t(r) = q_t(r \,|\, \mathbf{x}_1)$ for every $\mathbf{x}_1$ regardless of f_1, and condition (13) holds iff $t \in \Omega_\delta(p)$. Since d_t is constant over all $\mathbf{x}_1 \in I_1$, $\Omega_\delta(p)$ can be defined as

$$\Omega_\delta(p) = \{t \;:\; d_{\mathcal{T}}(t^\star, t) \le d_{\max}(p)\}, \qquad (14)$$

using the distance $d_{\mathcal{T}}$ from (2), where $d_{\max}(p)$ denotes the maximal distance $d_{\mathcal{T}}$ at which inequality (13) still holds. With this interpretation of $\Omega_\delta(p)$ being a ball of radius $d_{\max}(p)$, its volume increases monotonically with the radius, and therefore $V_\delta(p)$ attains a minimum at $p^\star$ iff $d_{\max}(p)$ does so.

[4] Technically, for the volume to be well-defined, the space of transformations has to be equipped with a measure, w.r.t. which the set $\Omega_\delta(p)$ is measurable. For general Lie groups, a natural choice is the Haar measure, which can be computed explicitly as the transformation of the volume form in the group parametrization domain. Tt can be shown that $\Omega_\delta(p)$ is a Borel set, and hence is measurable.

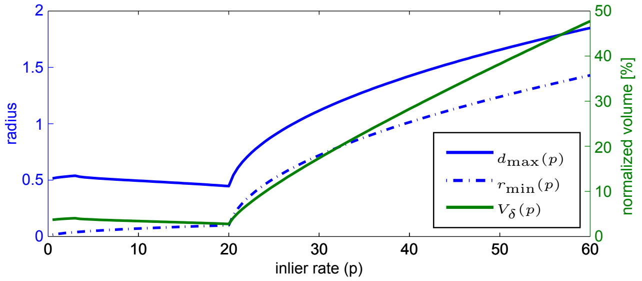

Figure 5. **Theoretical behavior of $d_{\max}(p)$, $r_{\min}(p)$ and $V_\delta(p)$ for 2D translations** for the case of $p^\star = 20\%$ (for a specific setting of $r^\star$ and δ). Dashed blue curve is the optimal error $r_{\min}(p)$, and the solid blue line shoes the distance $d_{\max}(p)$ obtained by solving (15). Green curve shows the volume $V_\delta(p)$. Left axis corresponds to the two blue curves; right axis corresponds to the green curve.

To show the existence of the latter minimum, we study the behavior of $d_{\max}(p)$ around $p^\star$. The probability $q_t(r_{\min}(p)+\delta \,|\, \mathbf{x}_1)$ from (13) can be expressed as

$$q_t(r_{\min}(p)+\delta \,|\, \mathbf{x}_1) = \dots \qquad (15)$$
$$\rho_{\text{in}} \cdot p^\star \cdot \psi(r^\star, r_{\min}(p)+\delta, d_{\max}(p)) + \dots$$
$$\rho_{\text{out}} \cdot (1-p^\star) \cdot \left[\pi(r_{\min}(p)+\delta)^2 - \psi(r^\star, r_{\min}(p)+\delta, d_{\max}(p))\right]$$

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 ψ). The value of $d_{\max}(p)$ can be 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 $p = p^\star$ (refer to Supplementary Materials for the full derivation), for a large range of values of $p^\star$ and $r^\star$, illustrated in Figure 4. □

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

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 bring 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.

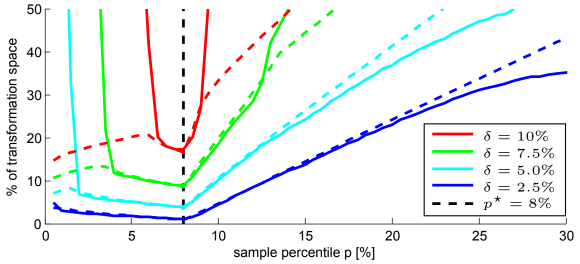

Figure 6. **Validating $V_\delta(p)$ in a synthetic 2D translation experiment.** We show the normalized volume $V_\delta(p)$ for different values of δ (shown in % of image size). A good match between theoretical (dashed) and empirical (solid) curves is observed in the vicinity of the true inlier rate $p^\star$. See text for details.

4.1. 2D translation - Synthetic data

We first wish to validate the formula for $V_\delta(p)$ 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 $V_\delta(p)$. Figure 6 shows results on a single instance of the problem generated according to our model, with inlier rate $p^\star = 8\%$ and inlier noise $r^\star$ of 5% of the size of I_2. We compare the volume $V_\delta(p)$ as obtained from empirical measurements (solid lines) to the theory (dashed lines) for several values of δ (color coded). There is an evident match between theory and practice at a certain interval around the true inlier rate $p^\star$ (black dashed line). The extent of this interval diminishes with the increase in δ, a phenomenon we discuss in the Supplementary Materials. In addition, note that all the curves attain a minimum at $p^\star$, as predicted by Proposition 1, even for high δ 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 $\mathcal{S}_\varepsilon$ of the space $\mathcal{T}$. The sampling density depends mainly on memory and time considerations, and tends to be effectively coarser for $\mathcal{T}$ 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 $\mathbf{r}_{\min}$ is an approximation of $r_{\min}(p)$; however, since $\mathcal{S}_\varepsilon$ is an ε-covering we incur an additive error of at most ε. Second, $\mathbf{v}_\varepsilon(p)/|\mathcal{S}_\varepsilon|$ approximates the normalized volume $V_\delta(\varepsilon)$. The fact that the sample is relatively uniform in the distance d_t ($\mathcal{S}_\varepsilon$ 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 previous 2D translation example. The first example is extreme

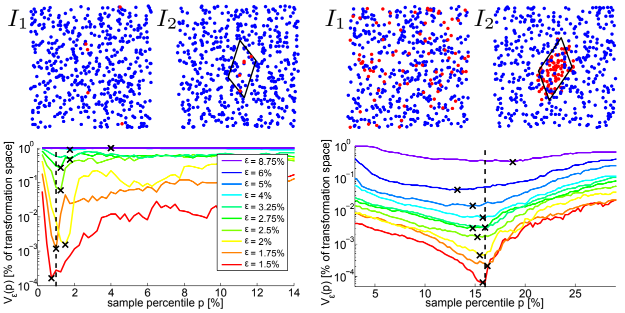

Figure 7. **Sensitivity to sampling of a synthetic 2D affine group example.** We show here results for two challenging examples. **Left example** (low inlier rate $p^\star$): $p^\star = 1\%$ and $r^\star = 1\%$. **Right example** (high inlier noise $r^\star$): $p^\star = 16\%$ and $r^\star = 8\%$. **In each example:** The image I_1 (top left) is mapped into image I_2 by an affine transformation (solid black parallelogram in top right). 500 matches were generated according to our model with the mentioned $p^\star$ and $r^\star$ (inlier matches are in red). We calculate $\mathbf{v}_\varepsilon$ using a sequence of step-sizes ε_i (color coded plots), and it is evident that the location of the minimum (black cross) stays roughly around $p^\star$.

in terms of the inlier-rate and the other one is extreme in terms of the inlier-noise. Each curve (color coded by sample density ε) shows our approximation $\mathbf{v}_\varepsilon(p)/|\mathcal{S}_\varepsilon|$ of the normalized volume $V_\delta(\varepsilon)$. 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 recomended 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.

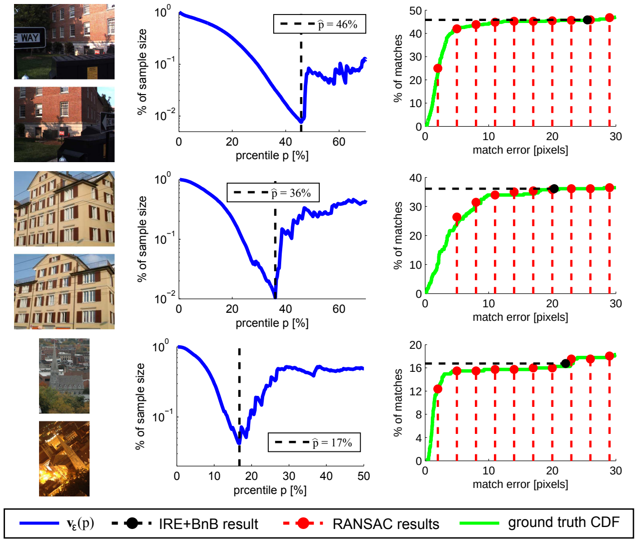

Figure 8. **Results on 2D homography with real data.** In each row we show **Left**: an image pair (from [16]); **Middle**: the prediction p^* (black dashed line) of Algorithm IRE as the minimal value of $\mathbf{v}_\varepsilon(p)$; **Right**: The result of Algorithm BnB (black circle) and the result of multiple USAC runs for different thresholds (red circles), shown against the CDF (green curve) of match-errors w.r.t. the ground-truth transformation. See text for further details.

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 $t_{\min}$, possibly providing a lower $r_{\min}(\hat{p})$ 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 $t_{\min}$ by reweighted least squares on the inliers. This heuristic may also lower $r_{\min}(\hat{p})$, 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 $\mathbf{v}_\varepsilon(p)$ 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

		Image pair				
sequence	method	1-2	1-3	1-4	1-5	1-6
bark	GMD	1.56	**3.45**	**2.53**	**1.14**	2.36
	USAC	**1.55**	3.49	**2.53**	1.15	**2.35**
graffiti	GMD	0.54	1.53	1.45	**6.55**	fail
	USAC	**0.53**	**0.85**	**0.98**	fail	fail
graffiti 4	GMD	**0.38**	**1.06**	**0.59**	**0.92**	**1.11**
	USAC	0.42	1.23	0.85	1.27	1.27
graffiti 5	GMD	0.75	**1.23**	**2.00**	**2.51**	**6.63**
	USAC	**0.62**	1.51	2.03	2.52	fail
wall	GMD	**1.24**	0.60	**1.29**	**1.56**	**2.12**
	USAC	**1.24**	**0.59**	1.33	1.66	2.81

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.

therefore USAC with such a threshold fails. Our method is less sensitive to the level inlier noise.

Table 1 compare the performance of our method (GMD) to USAC [16], in terms of Sampson error (see e.g. [6]), which compares to the ground truth transformation, over 5 viewpoint sequences from [14]. Both methods achieve similar accuracy, where USAC fails on the 3 most difficult pairs while we fail only on one. Our method is not claimed to be generally more accurate than USAC. Specifically in this experiment, error differences that are under one pixel fall below ground truth accuracy.

We now focus on the two hardest sequences (graffiti, graffiti-5) from [14]. Figure 9 shows a possible reason for the failures on these scenes. Specifically looking at graffiti 1-6, there seem to be virtually no correct matching SIFT features, and hence none of the methods is expected to work. Our method manages to solve the rest of the cases, despite the combination of high inlier noise and low inlier rate. For these same two sequences, Table 2 shows the estimated inlier rates and match errors attained by both methods. It is evident that USAC generally achieves lower match error as the noisiest inliers are discarded, a fact that may explain why it failed on extreme cases.

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 theoretic 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.

The assumption of uniform f_{in} and f_{out} was rather pragmatic to simplify the proof. Replacing the inlier (outlier) distributions would only affect the first (second) term of equation (15), where instead of a simple area calculation multiplied by the uniform noise ρ_{in} (ρ_{out}), a more complex integral would be used. For example, f_{in} could be replaced by the commonly assumed Gaussian noise.

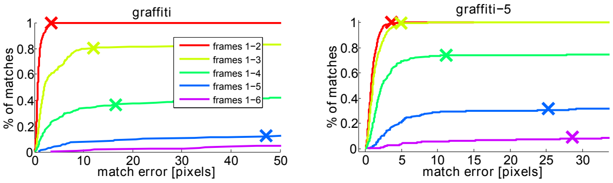

Figure 9. **Results on the two hardest sequences from [14].** Cumulative match error w.r.t the ground-truth are shown (color coded by pair number). As the view-point difference increases, the noise of inliers increases and their rate decreases. Our results shown by crosses in the respective colors.

The assumption of $\mathcal{T}$ being 2D translations is more fundamental, but we conjecture that it holds for a much wider range of transformation groups, as suggested by our experiments. Once this assumption is dropped, the proof must involve calculating the value of $p_t(r)$ in (6), which can be viewed as the expectation of $q_t(r|\mathbf{x}_1)$ over $\mathbf{x}_1$. The value $q_t(r|\mathbf{x}_1)$ depends in turn on the distance d_t, which now is not constant and depends on $\mathbf{x}_1$. This desired property alleviated the need to make any assumptions on the probability $f_1(\mathbf{x}_1)$ and made the "worst-case" analysis true for all points $\mathbf{x}_1$. Nevertheless, a tight bound could be achieved by a slightly more intricate "average-case" analysis, which we leave for future work.

		Image pair				
sequence	method	1-2	1-3	1-4	1-5	1-6
graffiti	GMD	99.8% 0.5±0.7	80.2% 1.8±1.8	36.6% 2.2±3.4	12.6% 8.5±14	fail
	USAC	87.8% 0.5±0.3	36.6% 0.7±0.3	19.5% 1.1±0.6	fail	fail
graffiti 5	GMD	99.8% 0.6±0.6	99.2% 0.9±0.8	74.2% 1.4±1.5	31.6% 3.0±4.9	9.3% 6.3±11
	USAC	89.2% 0.5±0.3	66.4% 0.6±0.3	33.6% 0.7±0.3	15.3% 1.5±0.7	fail

Table 2. **Inlier rates and match errors of homography estimation** on the graffiti and graffit-5 sequences from [14] using GMD and USAC. For each method, the detected inlier rate is shown in the first row, while the median±std of the inlier match errors is in the second. USAC achieves lower errors and std values since less matches are classified as inliers.

References

[1] J. Bazin, H. Li, I. S. Kweon, C. Demonceaux, P. Vasseur, and K. Ikeuchi. A branch-and-bound approach to correspondence and grouping problems. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(7):1565–1576, 2013. 2

[2] O. Chum and J. Matas. Matching with prosac-progressive sample consensus. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 220–226. IEEE, 2005. 2

[3] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. In *Pattern Recognition*, pages 236–243. Springer, 2003. 2, 7

[4] 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*, 24(6):381–395, 1981. 2

[5] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski. Non-rigid dense correspondence with applications for image enhancement. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2011)*, 30(4):70:1–70:9, 2011. 1

[6] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2, 7, 8

[7] S. Korman, R. Litman, S. Avidan, and A. Bronstein. Probably approximately symmetric: Fast rigid symmetry detection with global guarantees. *Computer Graphics Forum*, 2014. 3

[8] S. Korman, D. Reichman, G. Tsur, and S. Avidan. Fastmatch: Fast affine template matching. In *CVPR*, pages 2331–2338. IEEE, 2013. 1, 3

[9] H. Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1074–1080. IEEE, 2009. 2, 7

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 1

[11] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981. 1

[12] J. Matas and O. Chum. Randomized ransac with sequential probability ratio test. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1727–1732. IEEE, 2005. 2

[13] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1):63–86, 2004. 1

[14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630, 2005. 1, 7, 8

[15] C. Olsson, O. Enqvist, and F. Kahl. A polynomial-time bound for matching and registration with outliers. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. 2

[16] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J. Frahm. Usac: A universal framework for random sample consensus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):2022–2038, 2013. 2, 7, 8

[17] R. Raguram and J.-M. Frahm. Recon: Scale-adaptive robust estimation via residual consensus. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1299–1306. IEEE, 2011. 2

[18] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008. 7

[19] J. Yang, H. Li, and Y. Jia. Go-icp: Solving 3d registration efficiently and globally optimally. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1457–1464. IEEE, 2013. 2

[20] J. Yang, H. Li, and Y. Jia. Optimal essential matrix estimation via inlier-set maximization. In *Computer Vision–ECCV 2014*, pages 111–126. Springer, 2014. 2