Optimal Geometric Fitting Under the Truncated L_2-Norm

Erik Ask Olof Enqvist Fredrik Kahl
Centre for Mathematical Sciences, Lund University

Abstract

This paper is concerned with model fitting in the presence of noise and outliers. Previously it has been shown that the number of outliers can be minimized with polynomial complexity in the number of measurements. This paper improves on these results in two ways. First, it is shown that for a large class of problems, the statistically more desirable truncated L_2-norm can be optimized with the same complexity. Then, with the same methodology, it is shown how to transform multi-model fitting into a purely combinatorial problem—with worst-case complexity that is polynomial in the number of measurements, though exponential in the number of models.

We apply our framework to a series of hard registration and stitching problems demonstrating that the approach is not only of theoretical interest. It gives a practical method for simultaneously dealing with measurement noise and large amounts of outliers for fitting problems with low-dimensional models.

1. Introduction

Even though it is widely accepted that the truncated L_2-norm is a good way to model noise and outliers, its use has been hindered by the difficulty in solving the corresponding optimization problem. Local descent techniques are plagued by the many local minima often present and more sophisticated global optimization approaches suffer from the worst-case exponential time-complexity. In this paper, we show that the problem is in fact solvable in polynomial-time. This basic result serves as a basis for developing a practical and yet optimal approach to geometric fitting, simultaneously dealing with noise and outliers.

One motivating example is given in Figure 1. Here the objective is to register slices of prostate tissue stained in different colours. The staining shows the morphological characteristics of the tissue and is routinely used by uropathologists to diagnose cancer. As the images originate from different tissue slices, the local structure may look quite different in different images. Therefore, in order to obtain at least some correct correspondences, the acceptance threshold for

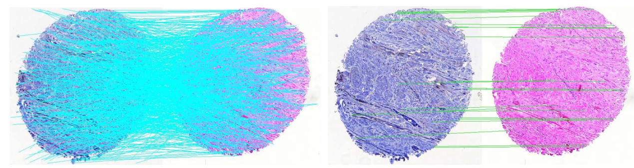

Figure 1. (Left) Images of two different stainings from a prostate biopsy with 791 hypothetical correspondences (cyan) obtained from SIFT. (Right) Same images, but now with 21 inlier correspondences (green) of the optimal truncated L_2-fit. The running time of our MATLAB implementation is 3s. See text for details.

the feature detector must be generously set. The downside is that this may produce a considerable amount of false correspondences. Note that the example in Figure 1 is the most difficult case in a database of 88 image pairs and more than 97% of the matches are outliers.

Related work. The most frequently used approach for dealing with outliers is based on RANSAC [10]. The basic assumption is that all minimal inlier-samples will produce an acceptable solution. However, it has been empirically observed that this is not valid in practice [12] and therefore numerous improvements of the basic technique have been proposed. For example, guided sampling methods have been proposed in [16, 6] and a strategy for local model refinement was developed in [12]. RANSAC and its variants have also been applied to multi-model estimation [18, 15]. Still, the fact remains that the estimator has no guarantee of finding the optimal solution.

Several recent works have focused on computing an optimal estimate based on branch-and-bound. For example, in [9, 8], robust estimators for 3D reconstruction is proposed and in [13], a formulation based on mixed integer programming is given. Similar ideas are presented in [3] for line clustering and vanishing point detection. These methods do not depend on initialization and converge to a global optimum. However, as they are based on branch-and-bound, the complexity of the algorithm is exponential.

In [14], it was shown that the problem of maximizing the number of inliers can be solved in polynomial-time in the number of measurements. This result was recently generalized in [7] by improving the complexity bound and weakening the conditions on the class of residual functions that can be handled. Our work is closely related at the conceptual level. Still, only the $0-1$ loss function (or more precisely, a piece-wise constant function) can be handled in [7], and not the statistically motivated truncated L_2-norm.

In the case of multiple models, the above methods (optimal or not) for estimating a single model can be applied sequentially. The sequential (or greedy) approach removes all correspondences that are deemed inliers for the most dominant model, and then the process is repeated. This, however, will not produce the optimal solution [18]. The approach may even produce "phantom" solutions [17]. Several heuristics have been proposed to overcome such artefacts, though there is no guarantee of optimality.

Contributions. The main contribution of this work is a proof as well as a practical algorithm that shows that the problem of estimating a model under the truncated L_2-norm can be achieved in polynomial time (in the number of measurements) given that there exists a method for solving the standard least-squares problem. Experimental results and comparisons to (exhaustive) RANSAC are given for challenging registration and stitching problems.

The second contribution concerns the problem of fitting multiple models. We show that maximizing the number of inliers can be done in polynomial-time given a fixed number of models and model parameters. Unlike greedy approaches that estimate one model at the time, we can guarantee that the optimal solution is obtained. The proof is based on similar ideas and concepts as in the truncated L_2-case.

2. Problem Formulation

Noise modelling. As noted above, most approaches for robust estimation in vision count the number of inliers to assess the quality of a solution, where inliers are - by definition - residual errors less than some prescribed threshold ϵ. This approach is simple and generally yields good results, but it does have its limitations. One problem is that the method might be sensitive to the choice of ϵ, but also that the distribution of the inlier errors is not modelled. In [4] a more refined loss function is proposed. The assumption is that inlier residuals have a clock-shaped error distribution similar to the Gaussian distribution, whereas outlier residuals have approximately uniformly distributed errors. These assumptions lead to the loss function

$$l(r) = -\log\left(c + \exp\left(-r^2\right)\right) \tag{1}$$

where r is the residual error; see Figure 2. It is also noted that a good approximation can be obtained by truncating the ordinary squared error.

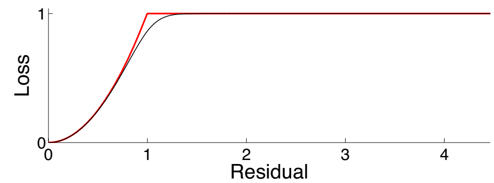

Figure 2. A robust loss function (solid black) as suggested in [4], and the truncated L_2-error (red) which can be optimized using the proposed framework; in this case truncated at $\epsilon = 1$.

A single model. We will work with a truncated quadratic loss function. For a residual r we compute the loss as

$$\ell(r) = \begin{cases} r^2 & \text{if } |r| \leq \epsilon, \\ \epsilon^2 & \text{otherwise.} \end{cases} \tag{2}$$

Problem 1. *Let D be a d-dimensional differentiable manifold, embedded in $\mathbb{R}^m$ ($m \geq d$) by a set of equality constraints $h_j(\theta) = 0$. Given a set of residual functions $r_i : D \to \mathbb{R}$, $i = 1, \ldots, n$, estimate a model $\theta \in D$ such that*

$$\sum_{i=1}^{n} \ell(r_i(\theta)) \tag{3}$$

is minimized.

Example. To make this more concrete, we look at the registration problem. Given two sets of points in $\mathbb{R}^2$, $\{x_i\}$ and $\{y_i\}$, we want to find

$$R = \begin{pmatrix} \theta_1 & -\theta_2 \\ \theta_2 & \theta_1 \end{pmatrix} \text{ and } t = \begin{pmatrix} \theta_3 \\ \theta_4 \end{pmatrix} \tag{4}$$

mapping one point set to the other. Here D is a 3-dimensional manifold that we embed in $\mathbb{R}^4$ using the equality constraint,

$$h(\theta) = \theta_1^2 + \theta_2^2 - 1 = 0. \tag{5}$$

The residuals are simply the Euclidean distances

$$r_i(\theta) = ||R(\theta)x_i + t(\theta) - y_i||. \tag{6}$$

Multiple models. In this case it is normally too difficult to work with the L_2-norm, so we settle with the $0-1$ loss

$$\ell(r) = \begin{cases} 0 & \text{if } |r| \leq \epsilon, \\ 1 & \text{otherwise.} \end{cases} \tag{7}$$

Problem 2. *Estimate a set of k models $\{\theta_1, \ldots, \theta_k\}$ with $\theta_j \in D$, such that*

$$\sum_{i=1}^{n} \min_{j=1,\ldots,k} \ell(r_i(\theta_j)) \tag{8}$$

is minimized.

This corresponds to minimizing the number of outliers.

A reformulation. Instead of solving our original problems directly, we will show that one can obtain the sought solutions by solving a number of simpler subproblems.

Definition 1. *Given two sets of residual functions I and O, let $D(I,O)$ denote the set of $\theta \in D$ such that $r_i(\theta) \leq \epsilon$ for all $r_i \in I$ and $r_i(\theta) \geq \epsilon$ for $r_i \in O$.*

Just as in [7], we introduce an dummy goal function f. In the end, f will be chosen to achieve simple equations, but now it is sufficient to say that it should be differentiable.

Problem 3. *Given two sets of residual functions I and O and a differentiable function $f : D \mapsto \mathbb{R}$,*

$$\min_{\theta \in D(I,O)} f(\theta). \tag{9}$$

The constraints on θ can also be written,

$$\begin{aligned} g_i(\theta) = r_i(\theta) - \epsilon \leq 0, &\ \textit{for all } i \in I \\ g_i(\theta) = \epsilon - r_i(\theta) \leq 0, &\ \textit{for all } i \in O \\ h_j(\theta) = 0, &\ \textit{for all } j\ , \end{aligned} \tag{10}$$

whith h_j as in condition 1. Note that all residuals r_i, $i = 1, \ldots, n$ need not be in $I \cup O$. In fact, we will only consider instances of Problem 3 with subsets. The *size* of the problem is defined to be the number of residuals $|I \cup O|$. We will show in the next two sections that both Problems 1 and 2 can be solved by considering instances of Problem 3 of size $\leq d$.

Definition 2. *An* FJ-point *to an instance of Problem 3 is a point that satisfies the Fritz-John conditions*[1] *for local optimality; see [2]. More precisely, a feasible point θ is an* FJ*-point if there is a non-trivial solution to*

$$\mu_0 \nabla f(\theta) + \sum \mu_i \nabla g_i(\theta) + \sum \lambda_j \nabla h_j(\theta) = 0 \tag{11}$$

with $\mu_i \geq 0$ and $\mu_i g_i(\theta) = 0$ for all i.

Naturally, we cannot solve any problem on this form. The following conditions will be sufficient for the theoretic discussion.

Condition 1. *We require that D is a d-dimensional differentiable manifold embedded in $\mathbb{R}^m$ using constraints $h_j(\theta) = 0$. We also require that the h_j are continuously differentiable and that the g_i's are differentiable.*

Condition 2. *If I is non-empty then $D(I,O)$ is bounded.*

[1]The Fritz-John conditions are closely related to the more well-known Karush-Kuhn-Tucker conditions.

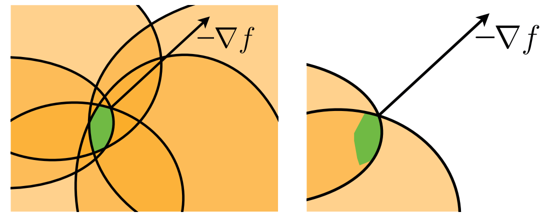

Figure 3. The residual functions trace out regions in D and for all θ in such a region, the inlier/outlier partition is constant. In order to find *one* point in the green region (left), it is enough to consider a subproblem with only $d = 2$ residuals (right).

Example. It is not hard to show that this condition holds for registration. Given a corresponding point pair (x_i, y_i), assume that the translation $|t| > |x_i| + |y_i| + \epsilon$. By the triangle inequality

$$|Rx_i + t - y_i| \geq |t| - |Rx_i| - |y_i| > \epsilon. \tag{12}$$

Hence the translation is bounded as long as there is at least one inlier, and as the rotation part has unit length this shows that $D(I,O)$ is bounded.

3. Fitting Under Truncated L_2-norm

For any feasible model point $\theta \in D$, the set of residuals will be partitioned into two sets, an inlier set I for which $r_i(\theta) \leq \epsilon$ for all $r_i \in I$ and the complement set of outliers O. The basic idea is to compute all such partitions - the main result is that there are only $\mathcal{O}(n^d)$ partitions - and then solve a standard least-squares problem for each inlier set. The result is illustrated in Figure 3.

If Condition 2 is satisfied and the residual functions are continuous, then there exists a minimizer to Problem 1. Due to lack of space, we omit the proof. The minimizer will have the following characteristic.

Lemma 1. *A minimizer θ^* to Problem 1 is also the global minimum to*

$$\min_{\theta} \sum_{r_i \in I^*} r_i^2(\theta), \tag{13}$$

where $I^* = \{r_i : r_i(\theta^*) \leq \epsilon\}$.

Proof. Assume to the contrary there exists a θ' such that

$$\sum_{r_i \in I^*} r_i^2(\theta^*) > \sum_{r_i \in I^*} r_i^2(\theta'). \tag{14}$$

Let O^* denote the set of residuals that are not in I^*. If we add $\epsilon^2|O^*|$ to the left hand side we get the total loss at θ^*, hence

$$\begin{aligned} \sum_{i=1}^{n} \ell(r_i(\theta^*)) &> \sum_{r_i \in I^*} r_i^2(\theta') + \epsilon^2 |O^*| \\ &\geq \sum_{r_i \in I^*} \ell(r_i(\theta')) + \sum_{r_i \in O^*} \ell(r_i(\theta')) = \sum_{i=1}^{n} \ell(r_i(\theta')) \end{aligned} \tag{15}$$

which is a contradiction. □

Lemma 1 shows that if we somehow knew the correct set I^*, we could find an optimal solution in truncated L_2 sense by solving a standard least squares problem. In many cases, this can be done efficiently. The key is then to find all candidates for the correct inlier set and the following theorem implies how this can be achieved.

Theorem 1. *Assume that Conditions 1 and 2 are satisfied. Let I be a non-empty subset of the residuals $\{r_i\}_{i=1}^n$ and O the remaining ones such that $D(I, O) \neq \emptyset$. Then at least one point in $D(I, O)$ is an* FJ*-point to an instance of Problem 3 of size $\leq d$.*

Lemma 2. *Let v, $e_1, \dots, e_n$ be elements in $\mathbb{R}^d$ $(n > d)$. Suppose*

$$\mu_0 v + \sum \mu_i e_i = 0, \tag{16}$$

where $\mu_i \geq 0$, not all zero. Then (possibly after renumbering), there exists $\bar{\mu}_i \geq 0$, not all zero, such that

$$\bar{\mu}_0 v + \sum_{i=1}^{d} \bar{\mu}_i e_i = 0. \tag{17}$$

Proof. According to the Theorem of Caratheodory, 0 can be written as a positive linear combination of no more than $d+1$ of the vectors. If at most d of the e_i's are used we are finished. It remains to consider the case when

$$0 = \sum_{i=1}^{d+1} b_i e_i \tag{18}$$

with all $b_i > 0$. If these e_i's do not span $\mathbb{R}^d$ they are in a subspace and we can use Caratheodory again to reduce the set. If they do span $\mathbb{R}^d$ then there exist c_i such that

$$v + \sum_{i=1}^{d+1} c_i e_i = 0. \tag{19}$$

Pick j such that c_j / b_j is minimized. Using (18) and (19),

$$b_j \cdot (19) - c_j \cdot (18) = b_j v + \sum_{i=1}^{d+1} (c_i b_j - c_j b_i) e_i = 0. \tag{20}$$

All coefficients here are ≥ 0 and the jth coefficient is 0. □

Proof (Theorem 1). Consider Problem 3 for this I and O. As $D(I, O)$ is closed and bounded, and f is continuous, a minimizer θ exists. By Theorem 4.3.2 in [2], θ will satisfy the Fritz-John conditions. More specifically,

$$\mu_0 \nabla f(\theta) + \sum \mu_i \nabla g_i(\theta) + \sum \lambda_j \nabla h_j(\theta) = 0, \tag{21}$$

where $\mu_i \geq 0$, $g_i(\theta) = r_i(\theta) - \epsilon$ are inequality constraints and $\mu_i g_i(\theta) = 0$, $i = 1, \dots, n$. As the h_j's are an embedding of a d-dimensional manifold in R^m, the set $\{\nabla h_j\}$ will span a $(m-d)$-dimensional subspace perpendicular to the manifold tangent space at θ. Let P be the projection operator onto this tangent space. By projecting (21) we get

$$\mu_0 P \nabla f(\theta) + \sum_{i=1}^{n} \mu_i P \nabla g_i(\theta) = 0. \tag{22}$$

Lemma 2 tells us that after renumbering, we can write

$$\bar{\mu}_0 P \nabla f(\theta) + \sum_{i=1}^{d} \bar{\mu}_i P \nabla g_i(\theta) = 0. \tag{23}$$

Now consider $\bar{\mu}_0 \nabla f(\theta) + \sum_{i=1}^{d} \bar{\mu}_i \nabla g_i(\theta)$. Clearly it must be perpendicular to the tangent space at θ and hence there exists $\bar{\lambda}_j$ such that

$$\bar{\mu}_0 \nabla f(\theta) + \sum_{i=1}^{d} \bar{\mu}_i \nabla g_i(\theta) + \sum \bar{\lambda}_j \nabla h_j(\theta) = 0. \tag{24}$$

It follows that θ is an FJ-point to an instance of Problem 3 of size $\leq d$. □

This theorem tells us that if we can compute all FJ-points for all instances of Problem 3 of size $|I \cup O| \leq d$, we can also find all possible partitions of inlier and outlier sets. For each FJ-point we simply check which residuals are $< \epsilon$ and $> \epsilon$, respectively. The only problem is residuals that are exactly ϵ. Generically, there will be at most d such residuals and hence we have 2^d possible inlier sets to examine. In practice, only one of these sets will make θ_{FJ} a real FJ-point satisfying the constraint $\mu_i \geq 0$ from (11).

Algorithm 1 Single-Model Fitting

For the $\mathcal{O}(n^d)$ instances of Problem 3 of size $\leq d$.
  Compute all FJ-points, denoted θ_{FJ}.
  For each θ_{FJ},
    Compute all $r_i(\theta_{FJ})$.
    Find the partitioning I, O induced by θ_{FJ}.
    Compute least squares solution θ^* for I.
    If this has the lowest loss so far, store θ^*.

Complexity. To actually find the FJ-points we need to study the Fritz-John conditions somewhat closer. It is clear that an FJ-point will have between 0 and d active inequality constraints. This leads to equations of the kind $g_i(\theta) = 0$. Note that these do not depend on the partitioning between I and O in Problem 3. Hence the number of equation systems that we need to solve to find all FJ-points is $\sum_{j=0}^{d} \binom{n}{j}$. Depending on the structure of the equations, each of these systems may have multiple solutions, but it does not depend on n. Finally, according to the discussion above each solution will normally only induce one partitioning, I, O of the residuals. For this I, we need to solve a least squares problem, which can often be done in $\mathcal{O}(n)$ time. Hence the total running time is $\mathcal{O}(n^{d+1})$ for Algorithm 1.

4. Multiple Models

The previous results also hold in the multi-model case, as fitting k d-dimensional models can be viewed as fitting one kd-dimensional model. However, the algorithm is less useful as it requires an efficient subroutine for computing the L_2-solution in the outlier-free case. This is in general not possible. If, on the other hand, we use the $0-1$ loss function, we can make the following important observation.

Theorem 2. *Assume that we to fit k models and that Conditions 1 and 2 are satisfied. An optimal solution to Problem 2 is a set of k FJ-points to instances of Problem 3 of size $\leq d$.*

Proof. As the goal function can only attain a finite number of values, a minimizer is guaranteed to exist. Consider such a minimizer, $\{\theta_1^*, \ldots, \theta_k^*\}$. We will study θ_1^* more closely, but note that the discussion holds for any θ_j^*. Let I_1^* and O_1^* be the set of inliers and outliers, respectively, to θ_1^*. Note that a residual can be an inlier to more than one model, but this will not matter. Consider the set $D(I_1^*, O_1^*)$ as in Definition 1. Clearly this set is non-empty since it contains θ_1^*. Moreover, for any θ in this set, $\{\theta, \theta_2^* \ldots, \theta_k^*\}$ is a solution to Problem 2. By Theorem 1 at least one point in $D(I_1^*, O_1^*)$ is an FJ-point to an instance of Problem 3 of size $\leq d$. This can be repeated for indices $2, \ldots, k$. □

The maximum k-cover problem. Assume that we have computed all FJ-points for instances of Problem 3 with $|I \cup O| \leq d$ residuals. The remaining problem is choosing k of these hypotheses such that Problem 2 is solved. Each hypothesis (or FJ-point) can be represented by its set of inliers I_i and we want to find a maximum k-cover,

$$\max_{|C|=k} \Big| \bigcup_{i \in C} I_i \Big|. \tag{25}$$

This is a well-known NP-hard problem, but it is easy to see that it is fixed-parameter tractable with respect to the number of models. More precisely, let H be the set of hypotheses, and assume that we want to fit k models. Since there are only $\binom{|H|}{k}$ possible choices, an exhaustive search can be done in polynomial time as long as k is fixed. Normally, a more efficient solution is to formulate the max k-cover problem as an integer linear program, and use standard solvers for this type of problem.

In summary, in order to optimally estimate a set of k models (Problem 2), one can use Algorithm 2. The worst-case running time is $\mathcal{O}(n^{kd+1})$.

5. Applications

We will look at three applications in more detail: Image registration, stitching and multi-model registration.

Algorithm 2 Multi-Model Fitting

For each instance of Problem 3 with size $\leq d$.
  Compute all FJ-points, denoted θ_{FJ}.
  For each θ_{FJ},
    Set $I_{FJ} = \{\, i \,|\, r_i(\theta_{FJ}) \leq \epsilon \,\}$.
Solve MAX k-COVER using the index sets $\{\, I_{FJ} \,\}$.

5.1. Image Registration

We return to the image registration problem introduced in Section 2. Given two sets of 2-vectors, $\{x_i\}$ and $\{y_i\}$, find a rotation R and a translation t that minimizes the truncated L_2-norm, as illustrated in Figure 1.

As per previous discussions we need to find all FJ-points to all instances of Problem 3 of size at most $d = 3$. We write the constraints g_i as

$$g_i(R, t) = (Rx_i + t - y_i)^T (Rx_i + t - y_i) - \epsilon^2. \tag{26}$$

We will need different polynomial solvers depending on the number of active constraints. With 3 active constraints we get 3 equations of type (26) and the embedding $h(\theta) = \theta_1^2 + \theta_2^2 - 1 = 0$. This amounts to 4 quadratic polynomial equations in 4 unknowns. Using methods presented in [5] we have implemented a minimal solver in MATLAB that runs in about 0.6 ms on a standard computer (Intel I5).

For problems of size < 3 we also need to consider the constraint posed on the gradients of g_i, h_j and f in (11). This constraint implies linear dependence of the gradients. These gradients are easy to calculate and using

$$\det \begin{pmatrix} \nabla f & \nabla g_1 & \nabla g_2 & \nabla h \end{pmatrix} = 0\,, \tag{27}$$

we obtain the necessary extra equations.

5.2. Image Stitching

We formulate the problem as Horn *et. al.* [11]. Given two sets of unit 3-vectors, $\{x_i\}$ and $\{y_i\}$, find a rotation R such that

$$\sum_{i=1}^{n} \ell(||Rx_i - y_i||) \tag{28}$$

is minimized. As before, $\ell(r) = \min\{r^2, \epsilon^2\}$. Unless the threshold ϵ is large, this is basically equivalent to minimizing squared angular errors.

The relevant inequality constraints for this problem are $g_i(R) = y_i^T R x_i + \epsilon^2/2 - 1 \geq 0$. The parameter space is 3-dimensional, one angle around each axis. We avoid trigonometry in the unknowns and embed the rotation matrix using a unit quaternion q. This does not alter the form of $g_i(R(q)) \geq 0$ but introduces an equality constraint $h(q) = ||q||^2 - 1 = 0$. Here $R(q)$ merely denotes a rotation matrix parametrized by the quaternion q.

For problems defined by 3 active constraints we again have a system of 4 equations in 4 unknowns. Using methods presented in [1] a polynomial solver exploiting the symmetry of the quaternion representation was constructed. For problems where only two inequality constraints are active we can no longer formulate a solvable system. Again we need (11) to obtain a fourth equation necessary to solve for the 4 unknowns in q. Just as for the case of registration we use the implied linear dependence and $\det(\nabla f \;\; \nabla g_1 \;\; \nabla g_2 \;\; \nabla h) = 0$ to obtain this equation. To simplify this expression we select $f(q) = q_1$, giving us $\nabla f = (1, 0, 0, 0)^T$ reducing the above determinant calculation to a subdeterminant calculation. Single active constraints can be handled similarly. Our MATLAB implementation of the polynomial solvers runs in about 10 ms.

5.3. Multiple Model Registration

As discussed in Section 4, we need to calculate, again, all FJ-points in where up to 3 constraints are active for single-model registration. This part of the scheme is identical to the single-model case. Each solution (or FJ-point) provides us with an inlier set. Picking the maximum k-cover among all possible inlier sets is then performed using CPLEX with the provided MATLAB interface.

6. Fast Outlier Rejection

It would naturally be a great advantage if outliers could be discarded early and not even considered as input to Algorithm 1. This section will present such a method for fast outlier rejection which works both for registration and stitching. Note that only correspondences that can be shown to not be part of an optimal inlier set will be discarded.

The technique iterates through the correspondences. For each correspondence c, we obtain a bound of the following type: *If c is an inlier, then the total loss l is larger than...* If this bound is higher than the best solution so far, we can remove c permanently from the discussion. Rather than working directly with the truncated L_2-norm we will use the following lower bound to the total loss, $\epsilon^2|O| \leq \sum_{i=1}^{n} l(r_i)$, where O is the minimal outlier set. Let us assume that residual i is below the threshold at optimum, that is, the ith correspondence is an inlier. Under this assumption, we will produce a bound on the optimal solution.

Proposition 1. *Suppose that for a set of corresponding points there exists a transformation T (for* registration *or* stitching*) such that all residuals are less than ϵ. Then there exists another transformation T' such that residual $r_i = 0$ and all other residuals r_j are less than 2ϵ, $r_j \leq 2\epsilon, j \neq i$.*

Proof. (*Registration*) Set $t' = y_i - Rx_i$ and $R' = R$. Then $||t - t'|| = ||t + Rx_i - y_i|| \leq \epsilon$ and hence for any j,
$r_j = ||R'x_j + t' - y_j|| \leq ||Rx_j + t - y_j|| + ||t - t'|| \leq 2\epsilon.$ □

Proof. (*Stitching*) Let $\alpha = \angle(y_i, Rx_i)$. Then a rotation R_α about $x_i \times y_i$ will map x_i exactly to y_i and $||R_\alpha|| \leq \epsilon$. So, set $R' = R_\alpha R$ and hence for any j,

$$\begin{aligned} ||R'x_j - y_j|| &\leq ||R_\alpha Rx_j - Rx_j|| + ||Rx_j - y_j|| \\ &\leq ||R_\alpha|| + \epsilon \leq 2\epsilon. \end{aligned}$$

□

This means that a bound for the number of 2ϵ-inliers given that $r_i = 0$ is also a bound for the number of ϵ-inliers given that $r_i \leq \epsilon$. For both registration and stitching, getting a bound is fairly easy as this constraint fixes the transformation up to a one-dimensional rotation using correspondence (x_i, y_i). We can parameterize this rotation with an angle α. Each of the remaining correspondences yields an interval constraint on this α (which should be interpreted modulo 2π). To get a bound on the number of inliers we need to find a point that lies in as many of these intervals as possible. This can be done by sorting the intervals and going through the sorted list. The computationally most costly part here is the sorting and hence the cost of this algorithm is $\mathcal{O}(n \log n)$. If we do this for each correspondence we get a cost of $\mathcal{O}(n^2 \log n)$, which is significantly cheaper than the optimal algorithm.

7. Experiments

We have performed a number of experiments on real data to demonstrate the validity of the approach on our three example applications. We compare our results with RANSAC, but in order for a fair comparison we use RANSAC with an exhaustive sampling of minimal subsets. The aim is to show that our approach is in fact practical, without having the drawbacks of a non-deterministic and non-optimal approach. For registration and stitching, the fast outlier rejection method of Section 6 is applied as a preprocessing step to Algorithm 1. Normally 90% to 98% of the outilers are eliminated in this step, but in rare examples this ratio drops and in the two worst examples only 0.2% and 11% are eliminated. For multi-model registration, we use Algorithm 2 with CPLEX for max k-cover, but before running the polynomial solver we do a simple test verifying that pairwise distances are consistent. This is a fast way to see if the solver will yield any real-valued solutions.

7.1. Registration

For the registration experiments the problem of matching two differently stained tissue slices of a prostate biopsy is addressed. Examples of such images are shown in Figure 4(a) out of a database of 88 pairs. These images vary greatly in the number of reliable matches, and can have high ratios of outliers (recall the example in Figure 1).

Truncation level is set at $\epsilon = 3$ pixels. Compared with standard RANSAC which optimizes the size of the inlier set

Figure 4. (a) Image pairs of prostate tissue in two different stainings. (b) Stitching example from FLICKR showing the view from the Eiffel Tower. This challenging example consists of two images taken under very different illumination conditions and resolutions, resulting in poor matching. (c) Image pair of a lung tissue biopsy which has three components. (Top) Hypothetical correspondences (cyan) obtained from SIFT. (Bottom) Inlier correspondences (green, blue and red) from the three estimated motions.

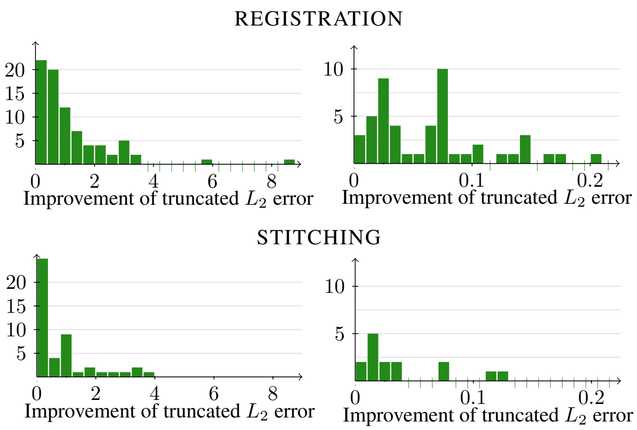

Figure 5. Histogram of performance differences between our optimal method and standard RANSAC (*left*) and RANSAC which evaluates the truncated L_2-error for each sample (*right*). The x-axis is scaled such that 1 unit corresponds to the cost of 1 outlier.

and in the end, computes a least-squares fit on this set, there were 8 pairs with no difference. The improvement for the other 80 pairs is shown in the upper, left of Figure 5. Compared with RANSAC that uses the same truncated quadratic loss, 42 pairs gave no difference. The improvement for the remaining 46 pairs is shown in the upper, right of Figure 5. The scaling of the x-axis is normalized by the squared threshold ϵ. This allows one to interpret the result as number of additional outliers. It is clear from the histograms that the optimal method significantly outperforms the best possible obtainable result from a standard inlier optimizing RANSAC approach. Even though the truncated L_2-RANSAC performs considerably better, it is still not optimal in half of the cases.

7.2. Stitching

A stitching example of a pair of images taken by different photographers at different times, downloaded from FLICKR with keywords "view from the Eiffel tower" is given in Figure 4(b). In total, a database of 53 image pairs were tested. The images were of size 640×480 and had very narrow overlap. The truncation was set to $\epsilon = 0.001$ corresponding roughly to 1.5 pixels. Compared with standard RANSAC, there were 5 pairs with no difference. The other 48 pairs are summarized in the lower, left of Figure 5. Similarly with the truncated L_2-RANSAC, there were 38 pairs with no difference while 15 pairs obtained improved results compared with the optimal method, see the lower, right of Figure 5.

The relative performance between our optimal method and RANSAC follows the same pattern as for the registration experiment. There is a significant difference compared to standard RANSAC while for the truncated L_2-RANSAC the performance difference is smaller. These findings are consistent with [12] where a truncated quadratic loss is also found to perform better than simply counting inliers in the RANSAC-loop.

	difference in # inliers										
# models	0	+1	+2	+3	+4	+5	+6	+7	+8	+9	+13
2	2	7	13	9	8	6	2	1	1	1	0
3	1	3	4	8	5	10	5	4	5	4	1

Table 1. The improvement of our method compared to sequential RANSAC. Note that RANSAC hardly ever finds the optimal solution even though the sampling of minimal sets is exhaustive.

7.3. Multiple Model Registration

For the experiments on simultaneous multiple model registration we emulate structures as the one displayed in Figure 4(c). The image shows a lung biopsy with three components, matched to adjacent layers of the same tissues. Due to the lack of availability of such data we use our prostate data to generate multi-model cases with close to the same number of inliers. For the cases with two models we used 20 inliers per model, and in total 40 outliers. For the three model case we used models of different dominance and set the inlier sizes to 10, 20 and 30. The number of outliers

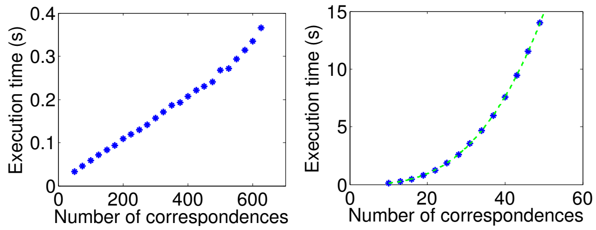

Figure 6. Timing experiments for registration. (Left) Runtime as a function of the number of correspondences for the fast rejection method. (Right) Runtime for Algorithm 1 as a function of problem size ($\approx$ the number of inliers). The dashed line is a $y = cx^3$-curve.

was set to match the total number of inliers. The results for both methods, compared to using a sequential RANSAC is displayed in Table 1. Average running time for the 2-model examples was 18 s and 90 s for fitting 3 models.

7.4. Runtimes and Time Complexity

By randomly selecting subproblems of different sizes from the registration data, we examined the running times of the fast outlier rejection and Algorithm 1 as a function of problem size. The results are shown in Figure 6. As most of the outliers are removed by the outlier rejection step and the runtime of the second step depends mainly on the number of inliers. This means that the total execution time of our algorithm varies from a couple of seconds to slightly over 3 h even if the number of input points is fixed. In practice though, there would be no need to run Algorithm 1 exhaustively in cases with hundreds of inliers.

The practical complexity for Algorithm 1 is cubic rather than the theoretical $\mathcal{O}(n^4)$. It shows that for this size problems the dominant cost is that of computing the FJ-points. The runtime of RANSAC is about 1 s for all experiments.

8. Conclusions

We have shown how to minimize the truncated quadratic loss function, which accurately models the noise for both inliers and outliers. Our experiments demonstrate that this yields a practical approach when combined with a fast outlier rejection step. One weakness is that for large number of inliers, it becomes infeasible to compute the optimal estimate. On the other hand, in such a case, it is not crucial to find all inliers in order to get an accurate estimate.

Our work opens up the possibility to develop new feature detectors. Current detectors are optimized to find a good inlier/outlier ratio, whereas our approach shows that it is possible to handle large amounts of outliers.

References

[1] E. Ask, Y. Kuang, and K Åström. Exploiting p-fold symmetries for faster polynomial equation solving. In *Int. Conf. Pattern Recognition*, Tsukuba, Japan, 2012. 6

[2] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: Theory and Algorithms*. Wiley, 1993. 3, 4

[3] J-C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys. Globally optimal line clustering and vanishing point estimation in manhattan world. In *Conf. Computer Vision and Pattern Recognition*, 2012. 1

[4] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge, USA, 1987. 2

[5] M. Byröd, K. Josephson, and K. Åström. Fast and stable polynomial equation solving and its application to computer vision. *Int. Journal of Computer Vision*, 2009. 5

[6] O. Chum and J. Matas. Optimal randomized ransac. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2008. 1

[7] O. Enqvist, E. Ask, F. Kahl, and K. Åström. Robust fitting for multiple view geometry. In *European Conf. on Computer Vision*, 2012. 1, 2, 3

[8] O. Enqvist, K. Josephson, and F. Kahl. Optimal correspondences from pairwise constraints. In *Int. Conf. Computer Vision*, 2009. 1

[9] O. Enqvist and F. Kahl. Robust optimal pose estimation. In *European Conf. on Computer Vision*, 2008. 1

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 1981. 1

[11] B. Horn, H. M. Hilden, and S. Negahdaripour. Closed-form solution of absolute orientation using ortonormal matrices. *Journal of the Optical Society of America A*, 1988. 5

[12] K. Lebeda, J. Matas, and O. Chum. Fixing the locally optimized ransac. In *British Machine Vision Conf.*, 2012. 1, 7

[13] H. Li. Consensus set maximization with guaranteed global optimality for robust geometry estimation. In *Int. Conf. Computer Vision*, 2009. 1

[14] C. Olsson, O. Enqvist, and F. Kahl. A polynomial-time bound for matching and registration with outliers. In *Conf. Computer Vision and Pattern Recognition*, 2008. 1

[15] R. Toldo and A. Fusiello. Robust multiple structures estimation with J-linkage. In *European Conf. on Computer Vision*, 2008. 1

[16] B. Tordoff and D. W. Murray. Guided sampling and consensus for motion estimation. In *European Conf. on Computer Vision*, 2002. 1

[17] J. Wills, S. Agarwal, and S. Belongie. A feature-based approach for dense segmentation and estimation of large disparity motion. *Int. Journal of Computer Vision*, 2006. 2

[18] M. Zuliani, C.S. Kenney, and B.S. Manjunath. The multiRANSAC algorithm and its application to detect planar homographies. In *Int. Conf. Image Processing*, 2005. 1, 2