Image and Vision Computing 52 (2016) 114–124

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Efficient algorithms for robust estimation of relative translation☆

Johan Fredriksson[a,*], Viktor Larsson[a], Carl Olsson[a], Olof Enqvist[b], Fredrik Kahl[a,b]

[a] Lund University, Sweden
[b] Chalmers University of Technology, Sweden

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form 19 February 2016
Accepted 2 May 2016
Available online 4 June 2016

Keywords:
Structure from motion
Epipolar geometry
Branch and bound
Two-view geometry

ABSTRACT

One of the key challenges for structure from motion systems in order to make them robust to failure is the ability to handle outliers among the correspondences. In this paper we present two new algorithms that find the optimal solution in the presence of outliers when the camera undergoes a pure translation. The first algorithm has polynomial-time computational complexity, independently of the amount of outliers. The second algorithm does not offer such a theoretical complexity guarantee, but we demonstrate that it is magnitudes faster in practice. No random sampling approaches such as RANSAC are guaranteed to find an optimal solution, while our two methods do. We evaluate and compare the algorithms both on synthetic and real experiments. We also embed the algorithms in a larger system, where we optimize for the rotation angle as well (the rotation axis is measured by other means). The experiments show that for problems with a large amount of outliers, the RANSAC estimates may deteriorate compared to our optimal methods.

1. Introduction

Two-view relative pose estimation is one of the cornerstone problems in computer vision. Many structure from motion pipelines solve it as a subproblem in order to build an overall reconstruction e.g. [2, 12, 25, 37]. Perhaps the most well-known method is the eight point algorithm by Longuet-Higgins [23], which gained popularity due to its simplicity and efficiency.

In this paper, we study the special case of translation estimation between two views. Either the camera is known not to rotate or the relative rotation can be estimated by other means. There are many situations where the camera undergoes pure translation, for example, in robotic applications. In addition, known rotation formulations have been shown to be useful subproblems in the structure from motion pipeline. The solution to the problem turns out to be very simple and computationally tractable and can be utilized as a subproblem when computing full 5 degrees of freedom relative pose. In [17], a branch and bound framework is introduced that searches over all rotations. As a subroutine, the translation of the camera is estimated using linear programming. It is assumed that no outliers are present in the data. The linear program could in principle be replaced with our approach to obtain an algorithm that works in the presence of outliers. Two-view geometry estimation is also used in image retrieval and reconstruction [3, 33, 35].

Known rotation formulations were first addressed in an algebraic framework in [34]. Later [19] showed that optimization of reprojection errors under the max-norm is a quasiconvex problem that can be solved using the bisection algorithm. Olsson and Kahl [30] used pseudoconvexity to design more efficient algorithms and Agarwal et al. [1] employed Gugat's algorithm. However, these works assume that there are no outlier correspondences. In fact, the max-norm's sensitivity to outliers makes these approaches highly sensitive to mismatches. Robust extensions to settings with outliers have been proposed [20, 29, 36, 40]. These offer very little guarantees in terms of optimality and typically handle only moderate amounts of outliers.

The main focus of this paper is estimation and outlier rejection in the presence of large amounts of outliers. For multiple view geometry problems, the predominant approach for dealing with outliers is RANSAC [13, 18, 27]. The method has been applied to related geometry problems such as relative pose estimation with rotation around one axis [14, 26]. Their formulation is motivated by the fact that many modern cell phones are able to directly measure the gravitation vector. A disadvantage with RANSAC methods is the randomness of the results and that the solution space is restricted to the hypotheses given by minimal subsets of the data. Furthermore, in settings with a large amount of mismatches this approach becomes costly because of the need to evaluate a large number of random

☆ This paper has been recommended for acceptance by Jan-Michael Frahm, Dr.-Ing.
* Corresponding author.
E-mail addresses: johanf@maths.lth.se (J. Fredriksson), viktorl@maths.lth.se (V. Larsson), calle@maths.lth.se (C. Olsson), olof.enqvist@chalmers.se (O. Enqvist), fredrik.kahl@chalmers.se (F. Kahl).

http://dx.doi.org/10.1016/j.imavis.2016.05.011
0262-8856/

samples. The approach offers no optimality guarantees meaning that even if there is a good solution the algorithm is not guaranteed to find it. Pham et al. [32] show that using hypothesis sets larger than the minimal number of samples generally improves the estimation. On the other hand this reduces the probability of randomly sampling a set of outlier-free measurements. In [28] it is shown that for pseudoconvex residuals (under the max-norm) exhaustive search over subsets of size $d+1$, where d are the degrees of freedom of the problem, is enough to find the globally optimal solution. These results are extended to non-convex problems in [10]. A direct application of the theory proposed in [10] to the two view relative translation problem would yield an $\mathcal{O}(n^3)$-algorithm for minimizing the number of outliers. Our approach is however significantly faster.

The branch and bound methodology has been applied to a variety of geometry problems including estimation in the presence of outliers. For quasiconvex residuals the properties of LP-type problems are exploited in [21], and the problem is solved using a tree search approach. This approach was further refined very recently in [7]. A related algorithm for dealing with outliers is given in [11]. In [22] a method for uncalibrated reconstruction based on an algebraic cost function and a mixed integer formulation is proposed. Similarly, in [4, 41], general outlier problems with linear residuals are addressed via mixed integer programming. While these methods guarantee optimality of the solution they are often based on computationally costly search schemes, with worst case exponential running time, resulting in prohibitively slow algorithms.

In this paper, we analyze in depth the case of two-view translation estimation. Our main contributions are as follows:

- We develop a provably optimal algorithm that, for a given error threshold ϵ, finds a translation which maximizes the number of correspondences with reprojection errors below ϵ. The worst case time complexity of the algorithm is $\mathcal{O}(n^2\log(n))$, where n is the number of hypothetical correspondences. Note that performing exhaustive search over all minimal subsets in a RANSAC-manner would result in an algorithm that has complexity $\mathcal{O}(n^3)$.
- We further develop a very fast branch and bound approach tailored for the two-view translation problem with the same optimality guarantees as the above algorithm. The key to designing an effective formulation is to find strong bounding functions that can be evaluated efficiently. We achieve this by a division of the parameter space S^2 using spherical triangles. Checking feasibility of a correspondence within a parameter triangle reduces to computing a few intersections between great circles, which can be done extremely efficiently using a simple cross-product.
- We experimentally demonstrate that our algorithms perform very well in several realistic and challenging set-ups including cases with known rotation, known rotation axis and small rotations. The baseline method for our comparisons is RANSAC and our experiments indicate that we not only obtain more accurate results, but also can be faster, particularly in presence of large amounts of outliers.

Preliminary versions of the algorithms have appeared in the conference papers [15, 16].

2. Background

In this work, we consider the spherical camera model which has previously been used in e.g. [17]. For spherical cameras the projections are formed by first applying a Euclidean transformation which takes the 3D-points to the camera's coordinate system followed by normalization. The projection of a 3D-point $\mathbf{X}$ in a spherical camera $(R, \mathbf{t})$ is formed as

$$\mathbf{v} = \frac{R(\mathbf{X}-\mathbf{t})}{\| R(\mathbf{X}-\mathbf{t}) \|}, \tag{1}$$

where R is a rotation matrix and $\mathbf{t}$ is a translation vector. This means that the image points are on the unit sphere S^2 in $\mathbb{R}^3$.

For the two-view translation estimation problem, we can w.l.o.g. assume that $R = I$ for both cameras by rotating the image points. Placing the first camera at the origin, the projections of a 3D-point $\mathbf{X}$ becomes

$$\mathbf{v} = \frac{\mathbf{X}}{\| \mathbf{X} \|} \quad \text{and} \quad \mathbf{v}' = \frac{\mathbf{X}-\mathbf{t}}{\| \mathbf{X}-\mathbf{t} \|}. \tag{2}$$

Given a correspondence pair $(\mathbf{v}, \mathbf{v}')$ in two images viewing the 3D point $\mathbf{X}$, the epipolar constraint tells us that the three vectors $\mathbf{v}$, $\mathbf{v}'$ and $\mathbf{t}$ should be coplanar as illustrated in Fig. 1.

The translation direction $\mathbf{t}$ from one camera center to the other can be assumed to have unit norm since the global scale cannot be determined in any case. In practice, one has to take into account measurement noise and ideally, we would like that the reprojections of $\mathbf{X}$ should lie as close to the measured image points $\mathbf{v}$ and $\mathbf{v}'$ as possible. We require that the reprojection is smaller than some predefined angular error $\epsilon \in \mathbb{R}^+$.

Definition 1. The corresponding point pair $(\mathbf{v}, \mathbf{v}') \in S^2 \times S^2$ is called an inlier for the translation $\mathbf{t}$ if there exists some $\mathbf{X} \in \mathbb{R}^3$ such that[1]

$$\angle(\mathbf{v}, \mathbf{X}) \leq \epsilon \quad \text{and} \quad \angle(\mathbf{v}', \mathbf{X}-\mathbf{t}) \leq \epsilon \tag{3}$$

for a given angular error $\epsilon \in \mathbb{R}^+$.

The above condition can equivalently be expressed as

$$\frac{\| \mathbf{v} \times \mathbf{X} \|}{\mathbf{v} \cdot \mathbf{X}} \leq \tan(\epsilon) \quad \text{and} \quad \frac{\| \mathbf{v}' \times (\mathbf{X}-\mathbf{t}) \|}{\mathbf{v}' \cdot (\mathbf{X}-\mathbf{t})} \leq \tan(\epsilon), \tag{4}$$

provided that the denominators are positive. For a fixed $\mathbf{t}$, each such condition constrains the 3D point $\mathbf{X}$ to lie in a convex cone, and hence the feasible set of 3D points satisfying Eq. (4) is convex and obtained as the intersection of two cones [19]. Next we state the problem we are interested in solving.

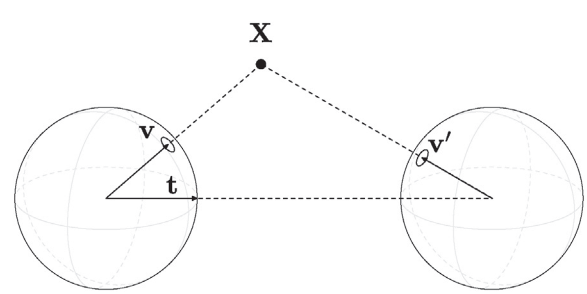

Fig. 1. Epipolar geometry. The translational direction $\mathbf{t}$, from one camera center to the other, should lie in the same plane as the two image points $\mathbf{v}$ and $\mathbf{v}'$ of the 3D point $\mathbf{X}$.