Robust Homography Estimation via Dual Principal Component Pursuit

Tianjiao Ding[1] Yunchen Yang[1] Zhihui Zhu[2] Daniel P. Robinson[3] René Vidal[4] Laurent Kneip[1] Manolis C. Tsakiris[1]

Abstract

We revisit robust estimation of homographies over point correspondences between two or three views, a fundamental problem in geometric vision. The analysis serves as a platform to support a rigorous investigation of Dual Principal Component Pursuit (DPCP) as a valid and powerful alternative to RANSAC for robust model fitting in multiple-view geometry. Homography fitting is cast as a robust nullspace estimation problem over either homographic or epipolar/trifocal embeddings. We prove that the nullspace of epipolar or trifocal embeddings in the homographic scenario, of dimension 3 *and* 6 *for two and three views respectively, is defined by unique, computable homographies. Experiments show that DPCP performs on par with USAC with local optimization, while requiring an order of magnitude less computing time, and it also outperforms a recent deep learning implementation for homography estimation.*

1. Introduction

A linear image-to-image map in homogeneous space—commonly known as a *homography* or a projective transformation in the plane—is a fundamental model in many contexts in photogrammetry and computer vision. Assuming the imaging device follows the pinhole or perspective camera model, the notion of homography provides a simple model of the image warping that occurs if the camera rotates and translates in front of a planar scene or undergoes a pure rotation. Thus, it has played an important role in camera calibration [1, 2], metric rectification [3, 4], panorama imaging [5], augmented reality [6, 7], optical flow based on piece-wise planar scene modelling [8] and video stabilization [9, 10]. The present paper focuses on robust homography estimation from image correspondences across two or three views, with points in each view extracted from a local invariant key point detector [11], and the correspondences established by appearance descriptor matching techniques.

The quality of the correspondences depends on a number of factors such as the invariance properties of the employed feature detector, motion blur, the texture of the environment, and the efficiency versus accuracy trade-off of the matching algorithm. It is therefore common that only a fraction of the identified cross-view correspondences are of high enough quality to yield an accurate homography estimate. As a result, the model fitting algorithm needs to be robust with respect to outliers in the set of correspondences, and the predominant scheme in computer vision to address this issue is RANSAC [12] and its more sophisticated derivatives developed over the last couple of decades, e.g., see [13, 14, 15, 16, 17] and references therein. Typically, these schemes may exhibit a high computational complexity due to a large ratio of outliers, high dimension of the inlier model, adaptive threshold determination, or local model refinement also known as local optimization.

On the other hand, over the past decade a variety of robust subspace learning methods have appeared, admitting efficient implementations and strong theoretical guarantees [18, 19]. Among these, Dual Principal Component Pursuit (DPCP) [20, 21, 22, 23, 24] stands as one of the few methods able to handle subspaces of high relative dimension or equivalently low relative codimension. This is precisely the case in homography estimation, since, under homographic embeddings, it can be cast as a robust hyperplane learning problem in $\mathbb{R}^9$ and $\mathbb{R}^{27}$ for two and three views respectively. DPCP formulates the hyperplane learning problem as an ℓ_1 non-convex optimization problem on the sphere, which can then be solved by a variety of methods such as recursive linear programming [20, 22], alternating minimization [25, 22], Iteratively-Reweighed-Least-Squares (DPCP-IRLS) [21, 22, 26], projected sub-gradient (DPCP-PSGM) [23, 24] and Riemannian sub-gradient (DPCP-RSGM) [27, 28] methods, or geodesic-gradient-descent [29]. In particular, DPCP-PSGM has been shown to tolerate $M < \mathcal{O}(N^2)$ outliers (here N, M is the number of inliers and outliers respectively), a notable result that contrasts sharply the typical guarantee $M < \mathcal{O}(N)$ of alternative methods [19][1]. More-

[1]School of Information Science and Technology, ShanghaiTech University, PRC [2]Department of Electrical and Computer Engineering, University of Denver, USA [3]Department of Industrial and Systems Engineering, Lehigh University, USA [4]Mathematical Institute for Data Science, Johns Hopkins University, USA. Correspondence to: Tianjiao Ding <dingtj@shanghaitech.edu.cn>, Manolis C. Tsakiris <mtsakiris@shanghaitech.edu.cn>

[1]A statistical analysis of this type of ℓ_1 non-convex optimization problem for subspace learning had already revealed that an arbitrarily large out-

over, DPCP-PSGM and DPCP-IRLS have been shown to be able to outperform RANSAC with the same running time budget and optimal threshold in surface normal [23, 24] and trifocal tensor [21] estimation.

In this paper we apply a group version of DPCP to the problem of robust homography estimation by casting the latter as a robust nullspace estimation problem. Here the groupings refer to the fact that each correspondence gives a group of embeddings. The nullspace of the matrix of embeddings is 1-dimensional under the traditional choice of homographic embeddings. On the other hand, it is also well understood that—for example in the case of a planar 3D point distribution—the epipolar or trifocal embeddings contribute a 3-dimensional or 6-dimensional nullspace, sitting inside 9-dimensional and 27-dimensional ambient space respectively. A fundamental property that we contribute here is that the structure of this nullspace is uniquely defined by the underlying homographies[2] (§2). The computation of these higher dimensional nullspaces can still be done by using extensions of the aforementioned methods from the unit sphere to the Stiefel manifold, the working choice here being DPCP-IRLS, and a global optimality theorem is given for the group DPCP problem (§3). Rigorous experiments suggest that DPCP performs competitively with the full-blown implementation of USAC [14], while running an order of magnitude faster (§4). When local optimization in USAC is disabled, DPCP not only remains faster, but is also significantly more accurate. Finally, DPCP outperforms the deep learning approach of [32] by a large margin.

2. Geometry: homographies and new insights

We start by giving a brief review of epipolar and homographic constraints for two views (§2.1). We discuss the traditional way of extracting homographies from 1-dimensional nullspaces of matrices of homographic embeddings. Then we show that homographies are also uniquely encoded in 3-dimensional nullspaces associated with epipolar embeddings. We then discuss the analogue in three views, namely homographic and trifocal constraints, and prove the unique recovery of homographies from 6-dimensional nullspaces of matrices of trifocal embeddings (§2.2). The importance of estimating homographies from higher-dimensional nullspaces is discussed in §3.

2.1. Two-view geometry

Epipolar embeddings. Let $P = [I\ 0]$, $P' = [A\ a]$ be the camera matrices for two perspective views $\mathcal{V}, \mathcal{V}'$. It is well known that if $x_i \leftrightarrow x_i'$ are the projections of points ξ_i in 3-space onto $\mathcal{V}$ and $\mathcal{V}'$, then $x_i'^\top F x_i = 0$, where $F := [a]A$ is the so-called fundamental matrix. This is the well-known epipolar constraint which we rewrite as

$$\phi(x_i, x_i')^\top \operatorname{vec}(F) = 0, \tag{1}$$

with the $\phi(x_i, x_i') \in \mathbb{R}^9$ bilinear functions of the correspondence pair x_i, x_i', referred to as epipolar embeddings. If the displacement between the two views $\mathcal{V}, \mathcal{V}'$ is non-degenerate and the 3D points ξ_i are in general position, then eight correspondences uniquely determine F up to scale. That is, the following 9×8 matrix has a 1-dimensional left nullspace spanned by $\operatorname{vec}(F)$:

$$[\phi(x_1, x_1') \ \cdots \ \phi(x_8, x_8')]$$

Homographic embeddings. Consider a plane π in 3-space whose normal vector is $(h^\top, 1)^\top \in \mathbb{R}^4$. Let ξ_i now be points on π and $x_i \leftrightarrow x_i'$ their projections onto $\mathcal{V}$ and $\mathcal{V}'$. Then $x_i' \sim H x_i$, where $H := A - a h^\top$ is the homography matrix. Here $x' \sim Hx$ means that x' and Hx are colinear, which implies that $[x_i']_\times H x_i = 0$.[3] Since the matrix $[x_i']_\times$ is of rank 2, out of these three linear equations in H, at most two are linearly independent. It is enough[4] to consider the first two which we write as

$$\psi_j(x_i, x_i')^\top \operatorname{vec}(H) = 0, \ \ j = 1, 2.$$

The $\psi_j(x_i, x_i') \in \mathbb{R}^9$ are homographic embeddings, which are bilinear functions of the correspondence pair x_i, x_i'. Assuming a non-degenerate configuration of the 3D points ξ_i on the plane π, 4 correspondences are sufficient to uniquely determine H up to scale, i.e., the left nullspace of the 9×8 matrix of homographic embeddings has dimension 1.

Homography estimation via epipolar embeddings. Now we discuss how to extract a homography from a nullspace of dimension higher than 1 arising from epipolar embeddings. We have not been able to find these arguments in the literature, especially those on the uniqueness of the homography.

Let H be a homography between $\mathcal{V}$ and $\mathcal{V}'$ induced by a plane π. It is well-known that the fundamental matrix F of $\mathcal{V}, \mathcal{V}'$ admits a factorization $F = [a]_\times H$. It is equally well-known that if one constructs the usual epipolar embeddings in $\mathbb{R}^9$ for 8 correspondences $x_i \leftrightarrow x_i' = Hx_i$, as in the 8-point algorithm, one will find a 3-dimensional nullspace. What is perhaps less well known is that the homography H uniquely defines the structure of this nullspace.

lier ratio can be tolerated provided that sufficiently many inliers are present [30, 31], an insight later confirmed by deterministic arguments in [22].

[2] One homography for two views, two for three views.

[3] Here for $a = (a_1, a_2, a_3)^\top \in \mathbb{R}^3$

$$[a]_\times = \begin{bmatrix} 0 & a_3 & -a_2 \\ -a_3 & 0 & a_1 \\ a_2 & -a_1 & 0 \end{bmatrix}$$

is the skew-symmetric matrix that represents the linear map $b \mapsto a \times b$ that takes vector b to its cross product $a \times b$ with a. Notice that $[a]_\times a = 0$.

[4] Since $(x_i')_3 \neq 0$ and $x_i'^\top [x_i']_\times = 0^\top$, one can verify that the third equation is dependent on the first two.

Proposition 1. *A homography represented by H induces a 3-dimensional vector space $\mathscr{F}_H = \{[v]_\times H : v \in \mathbb{R}^3\}$ of compatible fundamental matrices. Moreover, if $\mathscr{F}_{H'} \subseteq \mathscr{F}_H$ for some*[5] *$H' \neq 0$ then $\mathscr{F}_{H'} = \mathscr{F}_H$ and $H' \sim H$.*

Proof. $\mathscr{F}_H$ is a vector space since $[v_1]_\times + [v_2]_\times = [v_1 + v_2]_\times$ for any v_1, v_2. Now $\mathscr{F}_H$ is spanned by $[e_1]_\times H, [e_2]_\times H, [e_3]_\times H$ where the e_i's are canonical vectors. Since H is invertible these are linearly independent, thus $\dim \mathscr{F}_H = 3$. Let $F \in \mathscr{F}_H$. Then $F = [v]_\times H \in \mathscr{F}_H$ for some $v \neq 0$. Since H is invertible $[v]_\times H$ is a fundamental matrix. Moreover, it is compatible with the homography represented by H because $H^\top [v]_\times H$ is skew-symmetric.

For the last statement, we proceed in two steps. First we consider the case where H' is invertible. Let $x \in \mathcal{V}$ and let $x' \in \mathcal{V}'$ with $x' \sim Hx$. Let $F' = [v]_\times H' \in \mathscr{F}_{H'}$. If $\mathscr{F}_{H'} \subseteq \mathscr{F}_H$ then F' is a fundamental matrix compatible with the homography H. That is $x'^\top [v]_\times H' x = 0$. This is to say that $H'x \in \mathrm{Span}(x', v)$ for any $v \neq 0$ so that $H'x \in \mathrm{Span}(x')$. Since H' is invertible, $H'x \sim x'$ and therefore $H'x \sim Hx$. This is true for any x, thus $H' \sim H$.

Now we consider the general statement for any H'. The set of H''s such that $\mathscr{F}_{H'} \subseteq \mathscr{F}_H$ is a linear subspace $\mathcal{W}$ of $\mathbb{R}^{3\times 3}$. The set $\mathcal{U}$ of invertible 3×3 matrices is an open set in the Zariski topology of $\mathbb{R}^{3\times 3}$ [33, 34]. Then $\mathcal{W}_\mathcal{U} := \mathcal{W} \cap \mathcal{U}$ is an open set in the subspace topology of $\mathcal{W}$. We know that $\mathcal{W}_\mathcal{U}$ is non-empty because $H \in \mathcal{W}_\mathcal{U}$. Since non-empty Zariski open subsets of linear subspaces are dense, the Zariski closure of $\mathcal{W}_\mathcal{U}$ is $\mathcal{W}$ itself. But we proved above that $\mathcal{W}_\mathcal{U}$ is exactly the line spanned by H excluding the zero matrix. Adding back the zero matrix we get a 1-dimensional linear subspace of $\mathcal{W}$, which is a closed set. That is $\mathcal{W} = \mathcal{W}_\mathcal{U} \cup \{0_{3\times 3}\} = \{\lambda H : \lambda \in \mathbb{R}\}$. □

The vector space $\mathscr{F}_H$ of Proposition 1 is the 3-dimensional left nullspace $\mathcal{N}$ of the 9×6 matrix Φ of epipolar embeddings of 6 point correspondences $x_i \leftrightarrow Hx_i$. Proposition 1 asserts that H is uniquely encoded in any basis F_1, F_2, F_3 of $\mathscr{F}_H$ in the following strong form. Consider the homogeneous linear system in the c_{ij}'s and H'

$$\sum_{j=1}^{3} c_{ij} F_j = [e_i]_\times H', \quad i = 1, 2, 3. \tag{2}$$

To say that H' is a solution is to say that $\mathscr{F}_{H'} \subseteq \mathscr{F}_H$. Then Proposition 1 gives $H' \sim H$. In other words, (2) admits a unique up to scale solution in H', the homography H.

Even when the correspondences are slightly noisy, $\mathcal{N}$ need not have the structure of $\mathscr{F}_H$. One then computes an $\hat{H}$ such that $\mathscr{F}_{\hat{H}}$ is closest to $\mathcal{N}$ in a Euclidean sense as follows. Let $B_j \in \mathbb{R}^{3\times 3}$, $j = 1, \ldots, 6$ be the matrices that correspond to the top 6 left singular vectors of Φ. Following [35, A5.6],[36] $\hat{H}$ is obtained in closed form in terms of a singular value decomposition by solving

$$\min_{H'} \sum_{i,j} \|B_j^\top [e_j] H'\|_F^2 \text{ s.t. } \|([e_1]_\times\ [e_2]_\times\ [e_3]_\times) H'\|_F = 1 \tag{3}$$

2.2. Three-view geometry

Trifocal embeddings. The analogue of the fundamental matrix in three views is the (uncalibrated) trifocal tensor [35, 37, 38]. Consider three views $\mathcal{V}, \mathcal{V}', \mathcal{V}''$ with projection matrices $[I\ 0], [A\ a], [B\ b]$. Let $x \leftrightarrow x' \leftrightarrow x''$ be the projections on the three views of a point ξ in 3-space. In analogy with the epipolar constraint that captures the geometry of two views the triplet of correspondences satisfies

$$[x'] \Big(\sum_{i=1}^{3} x_i T_i \Big) [x''] = 0, \quad T_i := a_i b^\top - a b_i^\top, \tag{4}$$

where a_i, b_i are the ith columns of A, B and x_i is the ith coordinate of x. The matrices $T_i \in \mathbb{R}^{3\times 3}$ are the slices of the trifocal tensor $\mathcal{T}$. Conversely, any tensor (T_1', T_2', T_3') is a trifocal tensor if it admits a decomposition as in (4). Viewed as elements of $\mathbb{P}^{26}$ the trifocal tensors form a dense set in an irreducible projective variety of dimension 18.

The system (4) consists of 9 equations trilinear in x, x', x'' and linear in $\mathcal{T}$. Viewed as equations in $\mathcal{T}$, their coefficients are the trifocal embeddings of the correspondence $x \leftrightarrow x' \leftrightarrow x''$. Among these 9 trifocal embeddings only 4 are in general linearly independent. Then 7 general correspondences suffice to uniquely solve for $\mathcal{T}$ up to scale. This is done by computing the 1-dimensional left nullspace of a 27×28 matrix of trifocal embeddings.

Homographic embeddings. Suppose now that points ξ_i from a plane π are projected onto three views $\mathcal{V}, \mathcal{V}', \mathcal{V}''$ to yield correspondences $x_i \leftrightarrow x_i' \leftrightarrow x_i''$. Then there exist homographies H, G such that $Hx_i' \sim x_i$, $Gx_i'' \sim x_i$. As a consequence $\mathrm{rank}([x_i\ Hx_i'\ Gx_i'']) = 1$. Thus,

$$[x_i]_\times H x_i' = [x_i]_\times G x_i'' = [Hx_i']_\times G x_i'' = 0. \tag{5}$$

With $\mathcal{H}_{ijk} := (h_j \times g_k)_i$ the ith coordinate of the cross product of the jth column of H and the kth column of G, (5) can be written as [39]

$$\sum_{i,j=1}^{3} x_i x_j' \mathcal{H}_{ijk} = \sum_{i,j=1}^{3} x_i x_j'' \mathcal{H}_{ikj} = \sum_{i,j=1}^{3} x_i'' x_j' \mathcal{H}_{kji} = 0,$$

for $k = 1, 2, 3$. These are 9 equations linear in the homography tensor $\mathcal{H} \in \mathbb{R}^{3\times 3\times 3}$, which is a bilinear function of H, G and takes values in a codimension 10 irreducible algebraic variety of $\mathbb{P}^{26}$ [40, 41][6]. The coefficients are bilinear functions in pairs of correspondences and as in the

[5] H' is not required a-priori to be invertible.

[6] A naive argument is by counting the degrees of freedom of H, G.

case of two views can be thought of as homographic embeddings into $\mathbb{R}^{27}$. Among the 9 homographic embeddings produced by each corresponding triplet, 7 of them are in general linearly independent. Therefore, assuming a non-degenerate point configuration, 4 such triplets are sufficient to uniquely determine the 1-dimensional left nullspace of the 27×28 matrix of homographic embeddings, from which the homography tensor $\mathcal{H}$ is computed up to scale. H, G are subsequently obtained following the work [39].

Homography estimation via trifocal embeddings. Notably, two homographies can also be uniquely computed from a nullspace of dimension 6 arising from trifocal embeddings. Let π be a plane in 3-space. We have:

Proposition 2. *The homographies $\mathcal{V} \xrightarrow{H} \mathcal{V}', \mathcal{V} \xrightarrow{G} \mathcal{V}''$ induced by the plane π induce a 6-dimensional vector space $\mathscr{T}_{H,G}$ of trifocal tensors. Each $T \in \mathscr{T}_{H,G}$ is compatible with point correspondences $x \leftrightarrow x' \leftrightarrow x''$ obtained by projecting points from π. Moreover, $\mathscr{T}_{H,G}$ is uniquely determined by H, G and uniquely determines H, G up to scale.*

Proof. As per Proposition 1 H induces a 3-dimensional vector space $\mathscr{F}_H$ of fundamental matrices. These matrices have the form $[v]_\times H$ with v varying in $\mathbb{R}^3$. They induce a 3-parameter family of projection matrices $[H\ v]$ for $\mathcal{V}'$, all compatible with correspondences $x \leftrightarrow x'$ obtained by projecting points from π. A similar argument shows the existence of a 3-parameter family of projection matrices $[G\ d]$ for $\mathcal{V}''$ compatible with correspondences $x \leftrightarrow x''$ obtained by projecting points from π. As H, G are induced by the same plane, we have a 6-parameter family $[I\ 0]$, $[H\ v]$, $[G\ w]$ of camera projections all compatible with correspondenes $x \leftrightarrow x' \leftrightarrow x''$ obtained by projecting points from π. Each element of the family gives a trifocal tensor T with $T_i = h_i w^\top - v g_i^\top$, $i = 1, 2, 3$ where h_i, g_i are the ith columns of H, G. These form a 6-dimensional vector space $\mathscr{T}_{H,G}$. The last statement follows again from Proposition 1 since H uniquely determines $\mathscr{F}_H$ and is uniquely determined by it up to scale and similarly for $G, \mathscr{F}_G$. □

Let $x_i \leftrightarrow x_i' \leftrightarrow x_i''$ be correspondences obtained by projecting 6 general points from π. A consequence of Proposition 2 is that the 27×24 matrix of trifocal embeddings (4 embeddings for each correspondence) will have a 6-dimensional left nullspace. The homographies H, G can be uniquely recovered from that nullspace by means analogous to those for epipolar embeddings and similarly for the case of slightly imperfect correspondences.

3. Optimization: group-DPCP

Suppose we are given a set of N inlier groups $X = [X_1, \dots, X_N] \in \mathbb{R}^{D\times NK}$, where each group $X_i \in \mathbb{R}^{D\times K}$ contains K inlier points in its columns lying in a linear subspace $\mathcal{S} \subset \mathbb{R}^D$ of low relative codimension $c/D, c = D - \dim S$, contaminated by a set of M arbitrary outlier groups $O = [O_1, \dots, O_M] \in \mathbb{R}^{D\times MK}, O_j \in \mathbb{R}^{D\times K}$, and we wish to find $\mathcal{S}$. Let $\tilde{X} = [X\ O]\Pi \in \mathbb{R}^{D\times LK}$ with Π an unknown group permutation indicating that the segmentation of $\tilde{X}$ into inlier and outlier groups is not available a priori. With $\tilde{X}_i$ the i-th group of points of $\tilde{X}$ (could be inliers or outliers) and $1(y) = 1$ if $y \neq 0$ and $1(y) = 0$ otherwise, under mild general position assumptions on $\tilde{X}$, $\mathcal{S}$ is the orthogonal complement of the range-space of the unique global minimum of the optimization problem

$$\min_{B\in\mathbb{R}^{D\times c}} \sum_{i=1}^{L} 1\big(\|\tilde{X}_i^\top B\|_F\big) \quad \text{s.t.} \quad \operatorname{rank}(B) = c. \tag{6}$$

As (6) is computationally intractable and in practice the inlier groups do not lie exactly in $\mathcal{S}$, we replace it with the following robust ℓ_1 optimization problem on the Stiefel manifold (the set of orthonormal matrices)

$$\min_{B\in\mathbb{R}^{D\times c}} \sum_{i=1}^{L} \|\tilde{X}_i^\top B\|_F \quad \text{s.t.} \quad B^\top B = I_c. \tag{7}$$

The range-space of an optimal solution B^* of (7) contains c orthogonal directions of minimal ℓ_1 norm for the projected dataset and for this reason we refer to them as *dual principal components* in analogy with the classical principal components. In this work, we adopt the group-DPCP formulation (7) to address the robust homography estimation in the presence of a large number of mismatches[7]. For example, suppose we have a 2-view homography estimation problem with L given correspondences $\tilde{x}_i \leftrightarrow \tilde{x}_i'$, where N of them are high-quality $x \leftrightarrow x'$ and M of them are mismatches $o \leftrightarrow o'$. Recall from §2.1 that each correspondence $\tilde{x}_i \leftrightarrow \tilde{x}_i'$ contributes two homographic embeddings $\tilde{X}_i \in \mathbb{R}^{9\times 2}$, i.e., $K = 2$. The embeddings $X \in \mathbb{R}^{9\times 2N}$ of the high quality correspondences lie close to a hyperplane

Table 1: Two-view and three-view motion models, number K of embeddings per correspondence, embedding type, subspace type for homography estimation, and codimension parameter c to be used in (7).

Motion model	K	Embedding	Estimation of	c
2-view rigid body	1	epipolar	a hyperplane in $\mathbb{R}^9$	1
3-view rigid body	4	trifocal	a hyperplane in $\mathbb{R}^{27}$	1
2-view homography	1	epipolar	a 6-dim. subspace in $\mathbb{R}^9$	3
	2	homographic	a hyperplane in $\mathbb{R}^9$	1
3-view homography	4	trifocal	a 21-dim. subspace in $\mathbb{R}^{27}$	6
	7	homographic	a hyperplane in $\mathbb{R}^{27}$	1

[7] In the recent work [42] problem (7) was proposed for the optional local optimization step in RANSAC but the ℓ_1 norm was replaced by a Huber loss to deal with the non-differentiability of the ℓ_1 norm.

of $\mathbb{R}^9$ whose normal vector directly encodes the homography. On the other hand no linear structure is expected in the embeddings $O \in \mathbb{R}^{9\times 2M}$ of the mismatches. Then we may robustly estimate the homography by solving (7) with $c = 1$. Alternatively, we may work with epipolar embeddings where the task is to extract the homography from a 3-dimensional nullspace (§2.1). In that case we solve (7) for $c = 3$ (now $\tilde{X}_i$ are the epipolar embeddings) to get an orthonormal basis for that nullspace and subsequently extract the homography by solving (3). Similarly, a choice of $c = 1$ is needed for homographic embeddings of 3-view correspondences or $c = 6$ for trifocal ones (Table 1).

The importance of working with epipolar or trifocal embeddings in this context is computational: Note that the matrix of epipolar/trifocal embeddings has a smaller size than that of homographic embeddings. For example, each correspondence gives 1 epipolar embedding as opposed to 2 homographic ones in $\mathbb{R}^9$ or 4 versus 7 in $\mathbb{R}^{27}$ for the 3-view case. This results in faster computation and lighter memory usage, which is also evidenced in §4. Another potential advantage comes from the theory of robust subspace learning [18]: it is well-known that detecting outliers is a much simpler task when the intrinsic dimension d of the inlier data is small compared to the ambient dimension D. In such a regime a state-of-the-art toolbox of low-rank and sparse representation based methods with efficient implementations is available [43, 44, 45, 46, 47, 48, 49]. In contrast, the task is significantly more challenging when d/D is large [22, 50].

3.1. Global optimality of (7)

We present a novel condition for the global optimality of (7) in the noiseless case and otherwise under great generality. This condition is derived by extending the analysis of [23] for constraints of the form $B^TB = I_c$ as well as groupings of points. Suppose that for any $\{i_1, \dots, i_s\} \subset \{1, \dots, M\}$ we have $\operatorname{rank}([O_{i_1} \ \cdots \ O_{i_s}]) = \min\{D, Ks\}$. Any random O_i satisfies this with probability 1 and so do the homographic embeddings of random mismatches. We define three quantities that depend on either inliers X or outliers O:

$$\pi_X := \frac{1}{N} \min_{b\in\mathcal{S}\cap\mathbb{S}^{D-1}} \sum_{i=1}^{N} \|X_i^\top b\|_2$$

$$\rho_O := \frac{1}{M} \max_{B^TB=I_c} \left\|(I_D - BB^\top)\sum_{i=1}^{M} O_i \operatorname{Sign}(O_i^\top B)\right\|_F$$

$$\delta_O := \frac{1}{M}\Big(\max_{B^TB=I_c} \sum_{i=1}^{M} \|O_i^\top B\|_F - \min_{B^TB=I_c} \sum_{i=1}^{M} \|O_i^\top B\|_F\Big).$$

Here $\mathbb{S}^{D-1}$ denotes the unit sphere while for a matrix A we define $\operatorname{Sign}(A)$ as $A/\|A\|_F$ if $A \neq 0$ and 0 otherwise. These quantities are measures of uniformity of X, O. For example, the more uniformly the X_i are distributed in $\mathcal{S}$ the larger π_X is. Similarly, more uniformly distributed outliers O_i in $\mathbb{R}^D$ lead to smaller values for ρ_O and δ_O. Letting $\nu_O = \max_{i=1,\dots,M} \|O_i\|_F$, our result reads:

Proposition 3. *Any global solution $B^* \in \mathbb{R}^{D\times c}$ to (7) with $c = D - d$ is an orthonormal basis for $\mathcal{S}^\perp$ whenever*

$$\frac{M}{N} \frac{\sqrt{\left(\rho_O\sqrt{c} + \frac{\nu_O\sqrt{c}}{M}\min\left\{M, \frac{d}{K}\right\}\right)^2 + \delta_O^2}}{\pi_X} < 1. \quad (8)$$

To interpret (8) let us fix the ratio M/N to a constant value. Then the rest of the left-hand-side of (8) is a fraction whose denominator depends only on the inliers and its numerator only on the outliers. The more uniformly distributed X, O are the larger the denominator and the smaller the numerator become, i.e., (8) is more likely to be satisfied. A study of how (8) behaves when X, O are homographic embeddings is left to a longer version of this manuscript.

4. Experimental evaluation

We perform an experimental study of homography estimation using DPCP[8]. The algorithm used to solve (7) is DPCP-IRLS, to be henceforth referred to as DPCP(c) where c is the dimension of the target nullspace. For example, for 2-view homographic embeddings $c = 1$, while for epipolar embeddings $c = 3$. Similarly, for 3-view homographic or trifocal embeddings we have $c = 1, 6$ respectively. We employ a C++ implementation of DPCP for 2-view problems as well as a MATLAB implementation for 3-views.

We use three RANSAC variations as alternatives. The first one, USAC [14], is a fair representative of the state-of-the-art. USAC generates a model hypothesis using homographic embeddings from 4 sampled correspondences, and determines the proximity of a correspondence to the homography model via reprojection error. The threshold is set to 2 pixels across all experiments. Additional options are available such as local optimization, progressive sampling and model verification. We always use the latter, while (L) or (P) indicate whether the first two have been activated. An off-the-shelf integrated C++ implementation[9] of USAC is used, but is available only for 2-view problems. Thus we will compare this USAC implementation to the C++ implementation of DPCP for 2-views only.

For 3-view experiments we use our own optimized MATLAB implementation of two vanilla RANSAC algorithms termed RP-RANSAC and SD-RANSAC. These use reprojection and subspace distance errors respectively. The threshold for RP-RANSAC is consistently set to 2 pixels. This is used to optimally set a threshold for SD-RANSAC. The running time of both is set to be equal to the running time of the MATLAB implementation of DPCP.

[8] Experiments are run on a standard MacBook Pro 15 with a 6-core 2.2GHz processor and a total of 32GB memory.

[9] http://www.cs.unc.edu/~rraguram/usac/

Table 2: Homography estimation on 339 synthetically warped images from the MS-COCO dataset. Corner errors are in pixels and running times are in milliseconds.

	method	AUC	Corner Error	Time
mean	DPCP(1)	0.89	2.53	1.13
	DPCP(3)	0.88	5.08	0.88
	USAC	0.84	3.34	5.38
	USAC(L)	0.92	1.18	16.6
	USAC(P)	0.71	12.6	21.4
	USAC(PL)	0.84	6.75	22.7
median	DPCP(1)	0.94	0.81	0.98
	DPCP(3)	0.93	0.96	0.74
	USAC	0.88	2.37	3.28
	USAC(L)	0.95	0.48	13.0
	USAC(P)	0.76	8.58	1.43
	USAC(PL)	0.93	0.83	11.3

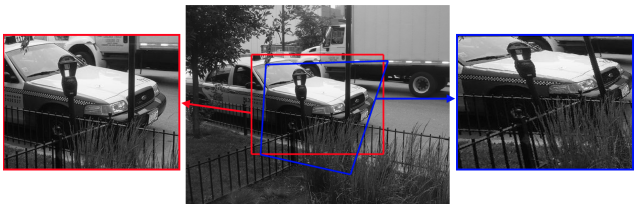


Figure 1: A random patch on the original image is taken as reference (red) and its corners are perturbed to get the second patch (blue). This defines a homography warp (right).

4.1. Homography warping

We follow the protocol of [51, 32] to generate warped images from the MS-COCO dataset [52] as shown in Fig. 1. We resize each image to 640×480 pixels in gray-scale and extract a 256×256 reference patch (red) at a random location. A perturbed patch (blue) is then generated by taking the corners of the reference patch and adding uniform noise up to 64 pixels. A ground-truth homography is computed by establishing correspondences between the corners of the two patches (§2.1). By warping the perturbed patch with the ground-truth homography, the second patch is brought back to the same size and format as the reference patch. The task is to estimate the homography given the two square patches. Correspondences are obtained by extracting and matching ORB features [53] between the two patches using a matching threshold of 0.3 and a ratio test threshold of 0.7.

Table 2 reports mean/median area under the precision-recall curve (AUC), corner error in pixels as defined in [51] and running time over 339 randomly sampled images from the MS-COCO dataset. To begin with, USAC(PL) achieves a mean and median corner error of 6.75 and 0.83 pixels respectively, which is in agreement with what was reported in [32, Figure 2]. Remarkably, DPCP outperforms both USAC(PL) and the deep learning approach reported in [32, Figure 2]: the latter has mean and median corner error of about 5 and 1.2 pixels while DPCP(1) has 2.53 and 0.81 respectively. On the other hand DPCP(3) appears to be less robust than DPCP(1) having twice the mean corner error of the latter. We conjecture that this is due to the homography estimation pipeline of DPCP(3), which involves two consecutive optimization problems, i.e. (7) followed by (3). It is the subject of current research to merge these into a single optimization problem, which is expected to improve the robustness. Note furthermore that the progressive sampling strategy of USAC seems to yield higher mean corner error, e.g., USAC(PL) has a mean error of 6.75 which drops to 1.18 for USAC(L), for which progressive sampling is switched off. Based on our investigations, the reason appears to be that the affinity score between matched ORB features—used for prioritizing the sampling [14, 54]—does not discriminate well between true and false matches.[10] Overall, USAC(L) gives the smallest mean corner error of about 1 pixel and the highest mean AUC value of 0.92. This is not a surprise since local optimization is known to lead to high accuracy [13, 14, 42]. However, note that DPCP(1) runs on average about 15 times faster than USAC(L) and has performance only slightly less accurate than USAC(L).

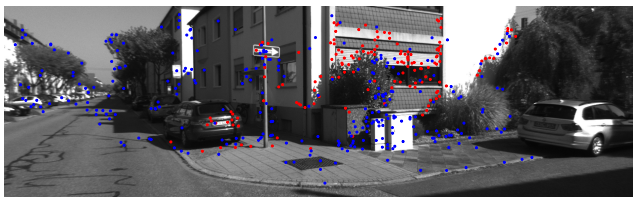
4.2. Structure from motion

We solve the structure from motion problem via homographies on sequences from the KITTI odometry [55] and the TUM RGB-D [56] datasets. The sequences cover outdoor driving scenes, as well as indoor man-made environments, which turn out to be well modeled by homographies. We match putative 2-view or 3-view correspondences over each pair/triplet of images.[11] We discard frames that have fewer than 30 correspondences among three views. For each method, homographies for view pairs $(\mathcal{V}, \mathcal{V}')$ and $(\mathcal{V}, \mathcal{V}'')$ are estimated. For the purpose of evaluation, we further convert the estimated homographies to calibrated homographies using the provided camera calibration information, and then perform homography decomposition [57, 58] to obtain rotation and translation matrices. For a homography, there are at most four pairs of possible rotations and translations, and the one with smallest error with regard to ground-truth is used for evaluation.[12] Finally, the 20 correspondences with the smallest reprojection error for each estimated homography model are used for refining the poses via bundle adjustment [59].

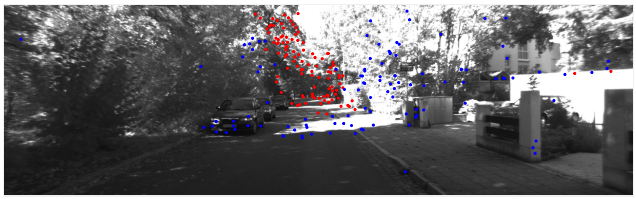
[10] We also explored other types of features such as SURF, and the conclusion was the same.

[11] We used SURF for the KITTI dataset and ORB for the TUM dataset.

[12] In practice, one could use, e.g., cheirality constraints to disambiguate. This complicates the evaluation step and hence it is not adopted here.



(a) Homography inliers (red) lie close to a wall plane.



(b) Homography inliers (red) are non-planar but have large depth and are close to the direction of motion.

Figure 2: A homography is fit to frames 101 and 105 from sequence 0 (see Figure 2a) and frames 3301 and 3305 from sequence 8 (see Figure 2b) of the KITTI dataset. Points compatible with the epipolar geometry (blue and red) and the homography model (red) are shown in the first view.

KITTI odometry. This is an autonomous driving dataset that includes images taken from a camera looking forward fixed on top of a car. As shown in Fig. 2 a homography can be successfully used due to the presence of correctly matched points that either 1) lie in a planar part of the scene (Fig. 2a) or 2) have large depth and are close to the direction of motion (Fig. 2b). For each sequence[13], frames $\{1, 101, 201...\}$ are sampled to be the first views, and then a gap of 4 is chosen to generate triplets of views, e.g., frames (100,105,109) will be a triplet for evaluation[14]. We compare the use of 2-view and 3-view embeddings by working with view triplets: in the former case we treat view triplets as pairs of two views.

Table 3 (top) reports mean rotation error and translation error in angles, and reprojection error for the ground-truth inliers before and after bundle adjustment (BA) for sequence 8 of KITTI. The lowest rotation error among the 2-view methods is $0.42°$ achieved by USAC(L), which takes 13.1ms to complete. Remarkably DPCP(1) converges in 0.64ms with a rotation error of only $0.1°$ higher than that of USAC(L). Moreover, DPCP(1) has the lowest translation error and the lowest reprojection error before and after BA: 1.46 pixels as opposed to 2.26 pixels for USAC(L).

Interestingly, lower errors are achieved by 3-view methods before BA. In particular, DPCP(1) gives the lowest rotation, translation and reprojection errors before BA among all methods. Moreover, DPCP(6) seems to be much more robust than DPCP(3) giving the second lowest errors. Notably, DPCP(6) converges in 2.8ms and has smaller errors than the USAC(L) for two views. This is interesting because—even though it is known that for the same number of correspondences optimizing over three views jointly gives more accurate poses than over pairs of two views independently [37, 60]—in practice there are fewer correspondences over 3-views. This experiment shows that working with 3-view embeddings can be advantageous.

Finally, even though for the purpose of homography estimation penalizing the reprojection error is more suitable than penalizing the subspace distance, SD-RANSAC appears to be more accurate than RP-RANSAC given the same time budget. This is because RP-RANSAC contains additional matrix multiplications and a matrix inversion in order to compute the symmetric transfer error. For example, in 4.76ms SD-RANSAC(1) completes 17 iterations as opposed to 8 iterations in 4.92ms for RP-RANSAC(1).

TUM RGB-D. This dataset includes sequences of different indoor environments taken from a hand-held RGB-D camera. We test the methods on two interesting sequences: 'fr3/nostructure_texture_near_withloop' where the scene is a planar wall with posters on it so that rich features can be detected and matched, and 'fr2/360_hemisphere' where the motion is almost translation free. For convenience we will refer to them as 'near' and 'hemisphere' respectively. We choose a frame gap of 20 for 'near' and 10 for 'hemisphere'.

As seen in the middle part of Table 3 the best performance for the sequence 'near' across all 2-view and 3-view methods is uniformly achieved by USAC(L). Rotation, translation and reprojection errors are $0.42°, 3.16°$ and 0.66 pixels respectively. Notice that USAC(L) takes 30.6ms to terminate. Remarkably, DPCP(1) for two views converges in only 0.84ms and has almost the same performance of $0.46°, 3.3°$ and 0.67 pixels. If we switch off local optimization, USAC may terminate slightly faster than DPCP(1), however at the cost of a significant degradation in terms of accuracy: it has more than $1°$ and $9°$ higher rotation and translation errors than DPCP(1). Note also that DPCP(3) performs almost as good as DPCP(1). Finally, the 3-view counterparts of these methods, that is DPCP(1) and DPCP(6), outperform their RANSAC analogues by a non-negligible margin. In particular, the angular translation error for DPCP(1) is $3.9°$ while RANSAC methods give at least $3°$ higher error.

The bottom of Table 3 indicates the results for the sequence 'hemisphere'. Note that all methods give very large translation errors owing to the fact that the motion in this sequence is almost purely rotational, which results in a low signal-to-noise ratio in the translation direction [61, 62]. Note that over three views, trifocal embeddings clearly produce better results than homographic ones.

[13] We tested sequences 0-10 which are shipped with ground truth poses, while in this paper we perform detailed study only of sequence 8.

[14] The fixed frame gap was manually chosen to ensure sufficient baseline and enough feature correspondences.

Table 3: Homography estimation from three views using 2-view or 3-view embeddings. Dataset/sequence in boldface. Rotation and translation errors are in degrees, reprojection error is in pixels and running time is in milliseconds. BA stands for bundle adjustment.

KITTI-8	before BA				after BA
2view-methods	Time	Rot.	Tran.	Repr.	Repr.
DPCP(1)	0.64	0.52	**4.02**	**2.02**	**1.46**
DPCP(3)	0.55	1.12	4.85	8.21	2.30
USAC	5.80	0.69	10.5	3.43	2.86
USAC(L)	13.1	**0.42**	6.04	2.39	2.26
	before BA				after BA
3view-methods	Time	Rot.	Tran.	Repr.	Repr.
DPCP(1)	4.63	**0.35**	**3.96**	**1.68**	1.73
DPCP(6)	2.80	0.42	4.11	1.81	**1.71**
SD-RANSAC(1)	4.76	0.50	5.77	2.33	2.19
SD-RANSAC(6)	2.89	0.55	7.32	2.35	2.15
RP-RANSAC(1)	4.92	0.80	7.82	5.64	11.3
RP-RANSAC(6)	3.41	1.69	10.7	36.5	9.75
TUM-near	before BA				after BA
2view-methods	Time	Rot.	Tran.	Repr.	Repr.
DPCP(1)	0.84	0.46	3.30	0.67	0.67
DPCP(3)	0.65	0.46	3.36	0.68	0.68
USAC	0.66	1.63	12.9	1.45	0.99
USAC(L)	30.6	**0.42**	**3.16**	**0.66**	**0.66**
	before BA				after BA
3view-methods	Time	Rot.	Tran.	Repr.	Repr.
DPCP(1)	7.64	**0.50**	**3.90**	**0.69**	**0.69**
DPCP(6)	4.77	0.51	3.96	0.69	0.69
SD-RANSAC(1)	7.97	0.94	6.93	0.97	0.82
SD-RANSAC(6)	5.07	0.88	7.57	1.20	0.86
RP-RANSAC(1)	8.33	4.45	23.0	36.4	7.91
RP-RANSAC(6)	6.86	6.01	29.5	48.0	37.1
TUM-hemisphere	before BA				after BA
2view-methods	Time	Rot.	Tran.	Repr.	Repr.
DPCP(1)	0.89	0.95	35.9	1.00	0.84
DPCP(3)	0.60	1.13	**34.7**	1.04	0.81
USAC	0.90	1.38	49.9	1.30	0.95
USAC(L)	27.8	**0.72**	37.6	**0.99**	**0.79**
	before BA				after BA
3view-methods	Time	Rot.	Tran.	Repr.	Repr.
DPCP(1)	4.12	2.46	36.5	2.69	1.01
DPCP(6)	2.70	**1.32**	**35.7**	**1.06**	**0.86**
SD-RANSAC(1)	4.28	1.85	47.1	1.67	1.82
SD-RANSAC(6)	2.82	1.47	47.6	1.17	0.97
RP-RANSAC(1)	4.44	5.09	53.5	36.9	37.4
RP-RANSAC(6)	3.41	5.31	54.5	12.9	4.63

Indeed, DPCP(6) gives the best performance among all 3-view methods with rotation and reprojection errors more than 1° and 1 pixel lower than those for DPCP(1). Moreover, DPCP(6) converges in about half the time required DPCP(1). In fact, across all experiments DPCP consistently converges faster with epipolar or trifocal embeddings as opposed to homographic embeddings. This is because, as already mentioned in §2 and §3, the former are more economic than the latter: for three views the data matrix of trifocal embeddings has $4/7$ times the size of the matrix of homographic embeddings, while this size ratio is $1/2$ for two-views. An additional potential factor is the effect of the lower intrinsic dimension of epipolar/trifocal embeddings; indeed as per Theorem 6[15] in [22] DPCP is known to converge faster for higher codimensions.

5. Discussion

We have shown that modern robust subspace fitting techniques are amenable to the solution of geometric vision problems. DPCP produces results that are competitive to state-of-the-art methods in terms of the achievable accuracy, and does so using substantially reduced computational effort. We have furthermore demonstrated that working over epipolar or trifocal embeddings can lead to another advantage in terms of computational efficiency, especially for the three-view scenario, at the cost of only slightly reduced accuracy. We believe the reduced accuracy is not inherent to our approach but rather an artifact of our computational pipeline that solves two consecutive optimization problems (efforts are being made to reduce this to a single problem). From a theoretical point of view, our contribution relies on a new global optimality condition for the DPCP algorithm. Moreover, in the context of homography fitting, we prove that the nullspace over epipolar or trifocal embeddings is uniquely defined by the underlying homographies, thus enabling a reduction in the intrinsic dimension and therefore another boost in computational efficiency.

We placed significant emphasis on comparing our proposed approach to the highly successful RANSAC algorithm through rigorous and exhaustive experimental evaluation. We applied all algorithms to multiple scenarios for both two and three view configurations while carefully considering practical implications such as a reduction in the number of correspondences over three views. Comparisons to state-of-the-art implementations of variants of RANSAC for regular CPU architectures were performed. From a practical point of view, we believe the available time budget is the most important concern in robust estimation, and in this regard we demonstrated that DPCP is superior to RANSAC.

[15] The upper bound on the number of iterations k^* given in Theorem 6 of [22] is a decreasing function of the codimension.

References

[1] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000. 1

[2] Z. Chuan, T. D. Long, Z. Feng, and D. Z. Li, "A planar homography estimation method for camera calibration," in *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, vol. 1, pp. 424–429, 2003. 1

[3] R. T. Collins and J. R. Beveridge, "Matching perspective views of coplanar structures using projective unwarping and similarity matching," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 240–245, 1993. 1

[4] D. Liebowitz and A. Zisserman, "Metric rectification for perspective images of planes," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 482–488, 1998. 1

[5] M. Brown and D. G. Lowe, "Recognising panoramas," *IEEE International Conference on Computer Vision*, vol. 2, pp. 1218–1225, 2003. 1

[6] G. Simon, A. W. Fitzgibbon, and A. Zisserman, "Markerless tracking using planar structures in the scene," in *IEEE and ACM International Symposium on Augmented Reality*, pp. 120–128, 2000. 1

[7] Z. Zhou, H. Jin, and Y. Ma, "Robust plane-based structure from motion," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1482–1489, 2012. 1

[8] J. Yang and H. Li, "Dense, accurate optical flow estimation with piecewise parametric model," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1019–1027, 2015. 1

[9] M. Grundmann, V. Kwatra, D. Castro, and I. Essa, "Calibration-free rolling shutter removal," *IEEE International Conference on Computational Photography*, 2012. 1

[10] Z. Zhou, H. Jin, and Y. Ma, "Plane-based content preserving warps for video stabilization," *IEEE Conference on Computer Vision and Pattern Recognition*, no. 1, pp. 2299–2306, 2013. 1

[11] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007. 1

[12] 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*, vol. 24, no. 6, pp. 381–395, 1981. 1

[13] O. Chum, J. Matas, and J. Kittler, "Locally optimized ransac," in *Joint Pattern Recognition Symposium*, pp. 236–243, Springer, 2003. 1, 6

[14] R. Raguram, O. Chum, M. Pollefeys, J. Matas, and J. M. Frahm, "USAC: A universal framework for random sample consensus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 2022–2038, 2013. 1, 2, 5, 6

[15] L. Moisan, P. Moulon, and P. Monasse, "Fundamental Matrix of a Stereo Pair, with A Contrario Elimination of Outliers," *Image Processing On Line*, vol. 5, pp. 89–113, 2016. 1

[16] D. Barath and J. Matas, "Graph-Cut RANSAC," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 6733–6741, 2018. 1

[17] D. Barath, J. Matas, and J. Noskova, "MAGSAC: Marginalizing sample consensus," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10189–10197, 2019. 1

[18] R. Vidal, Y. Ma, and S. S. Sastry, *Generalized Principal Component Analysis*, vol. 40. Springer Verlag, 2016. 1, 5

[19] G. Lerman and T. Maunu, "An Overview of Robust Subspace Recovery," *Proceedings of the IEEE*, vol. 106, pp. 1380–1410, 2018. 1

[20] M. C. Tsakiris and R. Vidal, "Dual principal component pursuit," in *ICCV Workshop on Robust Subspace Learning and Computer Vision*, pp. 10–18, 2015. 1

[21] M. C. Tsakiris and R. Vidal, "Hyperplane clustering via dual principal component pursuit," in *International Conference on Machine Learning*, 2017. 1, 2

[22] M. C. Tsakiris and R. Vidal, "Dual principal component pursuit," *Journal of Machine Learning Research*, vol. 19, no. 18, pp. 1–50, 2018. 1, 2, 5, 8

[23] Z. Zhu, Y. Wang, D. Robinson, D. Naiman, R. Vidal, and M. C. Tsakiris, "Dual principal component pursuit: Improved analysis and efficient algorithms," *Neural Information Processing Systems*, 2018. 1, 2, 5

[24] T. Ding, Z. Zhu, T. Ding, Y. Yang, D. Robinson, M. C. Tsakiris, and R. Vidal, "Noisy dual principal component pursuit," in *International Conference on Machine Learning*, pp. 1617–1625, 2019. 1, 2

[25] Q. Qu, J. Sun, and J. Wright, "Finding a sparse vector in a subspace: Linear sparsity using alternating directions," in *Advances in Neural Information Processing Systems 27*, pp. 3401–3409, 2014. 1

[26] G. Lerman and T. Maunu, "Fast, robust and non-convex subspace recovery," *Information and Inference: A Journal of the IMA*, vol. 7, no. 2, pp. 277–336, 2017. 1

[27] Z. Zhu, T. Ding, D. Robinson, M. C. Tsakiris, and R. Vidal, "A Linearly Convergent Method for Non-Smooth Non-Convex Optimization on the Grassmannian with Applications to Robust Subspace and Dictionary Learning," in *Advances in Neural Information Processing Systems 32*, pp. 9442–9452, 2019. 1

[28] X. Li, S. Chen, Z. Deng, Q. Qu, Z. Zhu, and A. M. C. So, "Weakly Convex Optimization over Stiefel Manifold Using Riemannian Subgradient-Type Methods," pp. 1–27, 2019. 1

[29] T. Maunu, T. Zhang, and G. Lerman, "A well-tempered landscape for non-convex robust subspace recovery," *Journal of Machine Learning Research*, vol. 20, pp. 1–59, 2019. 1

[30] G. Lerman and T. Zhang, "Robust recovery of multiple subspaces by geometric ℓ_p minimization," *Annals of Statistics*, vol. 39, no. 5, pp. 2686–2715, 2011. 2

[31] G. Lerman and T. Zhang, "ℓ_p-recovery of the most significant subspace among multiple subspaces with outliers," *Constructive Approximation*, vol. 40, no. 3, pp. 329–385, 2014. 2

[32] R. Ranftl and V. Koltun, "Deep Fundamental Matrix Estimation," in *European Conference on Computer Vision*, pp. 284–299, Springer Verlag, 2018. 2, 6

[33] J. Harris, *Algebraic geometry: a first course*, vol. 133. Springer Science & Business Media, 2013. 3

[34] D. Cox, J. Little, and D. OShea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2015. 3

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

[36] S. B. Heinrich and W. E. Snyder, "Robust Estimation of the Trifocal Tensor: A Comparative Performance Evaluation," *Preprint, https://shorturl.at/abhqu*, 2011. 3

[37] R. Vidal and R. Hartley, "Three-view multibody structure from motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 214–227, 2008. 3, 7

[38] A. Shashua and M. Werman, "Trilinearity of three perspective views and its associated tensor," in *IEEE International Conference on Computer Vision*, pp. 920–925, 1995. 3

[39] A. Shashua and L. Wolf, "Homography tensors: On algebraic entities that represent three views of static or moving planar points," in *European Conference on Computer Vision*, pp. 507–521, Springer, 2000. 3, 4

[40] C. Aholt and L. Oeding, "The ideal of the trifocal variety," *Arxiv preprint arXiv:1205.3776*, pp. 1–27, 2012. 3

[41] L. Oeding, "The quadrifocal variety," *Linear Algebra and Its Applications*, 2017. 3

[42] J. Šerỳch, J. Matas, and O. Drbohlav, "Fast l1-based ransac for homography estimation," in *21st Computer Vision Winter Workshop Luka Cehovin*, 2016. 4, 6

[43] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," *IEEE transactions on information theory*, vol. 58, no. 5, pp. 3047–3064, 2012. 5

[44] M. Soltanolkotabi, E. J. Candes, *et al.*, "A geometric analysis of subspace clustering with outliers," *The Annals of Statistics*, vol. 40, no. 4, pp. 2195–2238, 2012. 5

[45] M. Rahmani and G. K. Atia, "Coherence pursuit: Fast, simple, and robust principal component analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 23, pp. 6260–6275, 2017. 5

[46] C. Peng, C. Chen, Z. Kang, J. Li, and Q. Cheng, "RES-PCA: A scalable approach to recovering low-rank matrices," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 7309–7317, 2019. 5

[47] Y. Zhang, D. Shi, J. Gao, and D. Cheng, "Low-rank-sparse subspace representation for robust regression," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2972–2981, 2017. 5

[48] H. Q. Cai, J. F. Cai, and K. Wei, "Accelerated alternating projections for robust principal component analysis," *Journal of Machine Learning Research*, vol. 20, pp. 1–33, 2019. 5

[49] M. Rahmani and P. Li, "Outlier Detection and Robust PCA Using a Convex Measure of Innovation," *Neural Information Processing Systems*, pp. 1–11, 2019. 5

[50] Y. Wang, C. Dicle, M. Sznaier, and O. Camps, "Self Scaled Regularized Robust Regression," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3261–3269, 2015. 5

[51] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Deep Image Homography Estimation," *Arxiv preprint arXiv:1606.03798*, 2016. 6

[52] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*, pp. 740–755, Springer, 2014. 6

[53] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *IEEE International Conference on Computer Vision*, pp. 2564–2571, 2011. 6

[54] O. Chum and J. Matas, "Matching with PROSAC - Progressive sample consensus," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. I, no. I, pp. 220–226, 2005. 6

[55] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 6

[56] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of rgb-d slam systems," in *International Conference on Intelligent Robot Systems*, Oct. 2012. 6

[57] O. FAUGERAS and F. LUSTMAN, "Motion and Structure From Motion in a Piecewise Planar Environment," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 02, no. 03, pp. 485–508, 1988. 6

[58] Z. Zhang and a. R. Hanson, "3D Reconstruction Based on Homography Mapping," *ARPA Image Understanding Workshop*, pp. 249–6399, 1996. 6

[59] M. I. Lourakis and A. A. Argyros, "SBA: A software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software*, vol. 36, no. 1, 2009. 6

[60] P. Moulon, P. Monasse, and R. Marlet, "Global fusion of relative motions for robust, accurate and scalable structure from motion," *IEEE International Conference on Computer Vision*, pp. 3248–3255, 2013. 7

[61] R. Vidal and J. Oliensis, "Structure from planar motions with small baselines," in *European conference on computer vision*, pp. 383–398, 2002. 7

[62] O. Enqvist, F. Kahl, and C. Olsson, "Non-sequential structure from motion," in *IEEE International Conference on Computer Vision Workshops*, pp. 264–271, 2011. 7