Fast Two-View Motion Segmentation Using Christoffel Polynomials*

B. Ozbay[0000−0001−9964−6001], O. Camps[0000−0003−1945−9172], and M. Sznaier[0000−0003−4439−3988]

ECE Dept., Northeastern University, Boston, MA 02115, USA
ozbay.b@northeastern.edu, {camps,msznaier}@coe.neu.edu

Abstract. We address the problem of segmenting moving rigid objects based on two-view image correspondences under a perspective camera model. While this is a well understood problem, existing methods scale poorly with the number of correspondences. In this paper we propose a fast segmentation algorithm that scales linearly with the number of correspondences and show that on benchmark datasets it offers the best trade-off between error and computational time: it is at least one order of magnitude faster than the best method (with comparable or better accuracy), with the ratio growing up to three orders of magnitude for larger number of correspondences. We approach the problem from an algebraic perspective by exploiting the fact that all points belonging to a given object lie in the same quadratic surface. The proposed method is based on a characterization of each surface in terms of the Christoffel polynomial associated with the probability that a given point belongs to the surface. This allows for efficiently segmenting points "one surface at a time" in $\mathcal{O}$(number of points).

Keywords: Motion segmentation, Epipolar geometry, Algebraic clustering

1 Introduction

Motion segmentation –segmenting distinct moving objects in a sequence of frames– has a wide range of applications in computer vision and robotics [1–5].

While it is possible to perform trajectory association and object segmentation jointly [6, 7], this requires solving expensive optimization problems. Hence, most algorithms require feature correspondences between two or more frames to be given as input.

Most multi-frame approaches segment moving objects by clustering feature trajectories, under the assumption of an affine camera projection model. In this scenario, the trajectories lie in linear or affine subspaces and can be found by using subspace clustering [8–18], or factorization algorithms [19–23]. While these approaches perform well on benchmarks such as the Hopkins 155 dataset [24] they have several drawbacks. Since they rely on an affine projection model they do not perform well when images have perspective distortion. In addition, point trajectories require tracking features across multiple frames, which is more expensive and difficult than finding

* This work was supported in part by NSF grants IIS–1814631 and CNS–2038493, ONR grant N00014-21-1-2431 and U.S. DHS grant 22STESE00001-01-00.

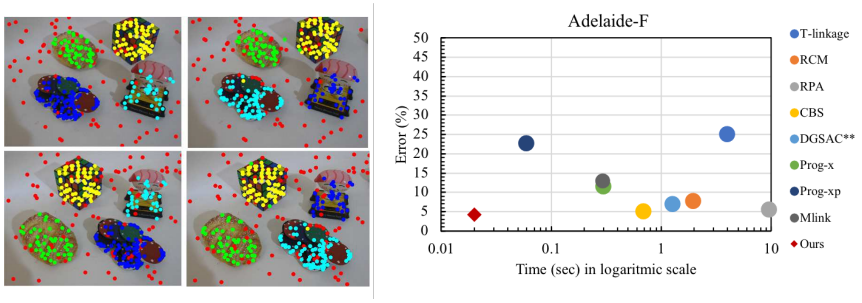


Fig. 1: Left: Sample motion segmentation results for the Adelaide-F data-set. Segmentation for two views of ground truth (1^{st} col.) and our algorithm (2^{nd} col.) each view in one row. Colors of feature points indicate their label where red shows outliers. Right: Comparison of time versus average segmentation error for the entire Adelaide-F data-set. The proposed method is both the most accurate (4.15% error) and fastest (0.02 secs per image pair).

singleton correspondences between a pair of frames. Finally, occlusions and missing features in intermediate frames often result in a relatively low number of reliable trajectories to work with. [25–28] handle perspective effects by combining results from two-view correspondences. However, they still require tracked trajectories. More recent approaches [29–32] avoid this requirement, for instance by using triplets of images [31]. However, despite achieving higher accuracy, trifocal tensors fail to classify a high percent of available pairs.

Two-view methods avoid the need for trajectory acquisition, by using the epipolar geometry between correspondences to cluster feature pairs. Thus, they can work with full perspective views. These approaches [33–38] relate pairs of corresponding features through geometric constraints, such as the Longuet-Higgins equation and enforce that features which belong to the same rigid object must be related by the same fundamental matrix. However, while it is easier to find a large number of correspondences between two frames than to obtain many long trajectories across multiple frames, these correspondences are often corrupted with outliers, making the segmentation task more difficult.

A popular approach to eliminate outliers while estimating a single fundamental matrix between two views of a static scene [39–44] is to use random sample consensus (RANSAC) [45]. RANSAC and related sampling methods have also been used when seeking multiple structures [46–48], including motion segmentation. However, the presence of multiple structures necessitates relatively expensive sampling in order to guarantee a given probability of achieving the correct segmentation. Other techniques [49–54] follow a preference-based approach, where the distribution of residuals of individual data points with respect to the models is inspected. However, the model step in these approaches suffers from low accuracy and depends on the bin size. Perhaps the closest approach in spirit to the one proposed in this paper is [12], where two-view motion segmentation is recast into a sparse subspace clustering form. While effective, this approach requires solving first a computationally expensive optimization, followed by a spectral clustering step. Hence, its computational complexity scales at least as $(\text{number of points})^3$.

In this paper we introduce a robust, computationally efficient algebraic approach for motion segmentation from feature correspondences between two perspective images (Fig. 1). The approach is based on a Christoffel polynomial characterization of the support set of the (unknown) probability distribution associated with data in each of the quadratic surfaces corresponding to points in the same object. This polynomial, which has low (high) values at inliers (outliers) to a given surface allows for efficiently segmenting the objects "one at a time, by identifying all points with high probability of belonging to the same object (Fig. 2). These points are removed from the population and the process is repeated until all points have been labeled. If desired, an outlier removal step can be implemented prior to starting the process, by considering the Christoffel polynomial corresponding to the joint distribution over all objects and identifying correspondences with high probability of being outliers. Notably, computing the Christoffel polynomial from the correspondence data involves the singular value decomposition (SVD) of a matrix whose size depends only on the number of objects, leading to an algorithm whose computational complexity scales linearly with the number of data points. These results are illustrated with standard datasets, showing that the proposed approach offers the best trade-off between error and computational time (Fig 1): it is at least one order of magnitude faster than the best method (with comparable or better accuracy), with the ratio growing up to three orders of magnitude for image pairs with larger number of correspondences. In terms of computational time, it is always the fastest, with an error rate at least 50% smaller than the runner up, with this number growing up to 3 times as the number of correspondences increases.

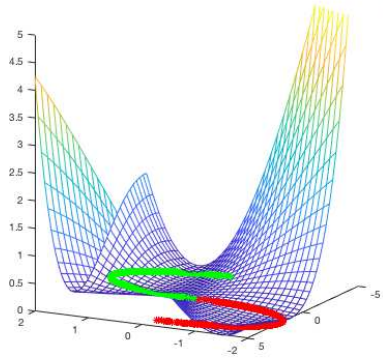


Fig. 2: One at a time segmentation of two second order curves. Since the green points are outliers to the red curve, the Christoffel polynomial of the latter has values higher than 0.5 at these points.

2 Related work

Two view motion segmentation methods can be roughly grouped into sampling based and model fitting based.

2.1 Sampling based approaches

Most of the sampling based methods build on RANSAC [45], which searches for consensus between randomly sampled minimal sets to estimate a single model. Some extensions of RANSAC, including [40, 42, 55], use a similar methodology on problems with multiple structures and depend on some form of prior knowledge such as matching scores, spatial distance or super-pixel information. Other variations [39, 56, 57], use the knowledge collected from previous steps to guide sampling in subsequent steps

and speed up model estimation. For a more comprehensive overview of variations of RANSAC please see [58, 59]. Note that although there are many advances in model estimation for both homography and fundamental matrix estimation [60–62] since these methods are tailored towards obtaining accurate models rather than accurate segmentation, they are not considered in the scope of this work.

2.2 Model fitting based approaches

The T-linkage approach [52] starts with random sampling to generate m hypothesis from minimum sample sets. Then, each data point is described by a *preference function* taking values between 0 and 1. Finally, a greedy bottom-up agglomerative clustering yields a partition of the data by merging points with similar preferences. Later in RPA [54] the authors exploited a robust M-estimator, combined with robust component analysis and non-negative matrix factorization. Some of the more recent preference based methods are MCT [63], HF [64], HOMF [65] and MultiLink [66]. MCT extends T-linkage to handle nested models such as planar sides of a cube whereas MultiLink extends preference representations to multiple, not necessarily nested, mixed classes of models. Although these model fitting based approaches generally achieve good performance on the fitting accuracy, they have a high computational complexity.

[53] learns a Random Cluster Models (RCM) to generate hypotheses using non-minimal subsets of samples. Point preferences are organized in a graph and then graph cuts are used to optimize the fitting. The MSHF method [67] clusters model hypotheses using hypergraphs constructed from data where each hypothesis indicates an instance in data. However, this method seeks modes from the generated model hypotheses which may lead to suboptimal fitting results, for cases where the generated model hypotheses do not contain all model instances. A similar method, Prog-x (P-x) [68], interleaves sampling and geometric multi-model fitting using a modified RANSAC to progressively explore the data through model proposal and optimization steps. Prog-xp (P-xp) [69] extends P-x by introducing a new problem formulation which allows points to be assigned to more than one model. These consensus analysis based methods can achieve a high fitting accuracy if provided by a high quality model hypothesis.

All the above approaches use random sampling to grow clusters and reject outliers. In contrast, our approach is deterministic and proceeds by selecting the most reliable data available at each step.

3 Notation

$\mathbf{A} \succeq 0$	matrix $\mathbf{A}$ is positive semidefinite.
$\mathscr{P}_n^d$	subspace of n^{th} degree homogeneous multivariate polynomials in d variables.
$s_{n,d} \doteq \binom{n+d-1}{d-1}$	number of monomials of degree n in d variables.
$\mathbf{v}_n(\mathbf{x}) \doteq$	$\begin{bmatrix} x_1^n & x_1^{n-1}x_2 & \dots & x_d^n \end{bmatrix}^T$ degree n Veronese map of $\mathbf{x} = \begin{pmatrix} x_1 & \dots & x_d \end{pmatrix}^T$
$\mathscr{E}_\mu(x)$	Expected value of x with respect to the probability density function μ.
$\lvert\mathscr{S}\rvert$	number of elements in the set $\mathscr{S}$.

4 Problem setup

The goal of this paper is to assign correspondences to objects. As discussed below, this problem is equivalent to algebraic variety clustering, where points need to be assigned to a known number of unknown second order varieties. Consider a set of N (inlier) correspondences $\mathscr{C} = \{(\mathbf{x}_1, \mathbf{x}_2)_i, i = 1, \ldots, N\}$ between two perspective views of a scene with a known number M of rigid objects. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are corresponding features that belong to object j, they must satisfy the epipolar constraint:

$$\mathbf{x}_1^T \mathbf{F}^{(j)} \mathbf{x}_2 = 0 \tag{1}$$

where $\mathbf{F}^{(j)} \in \mathbb{R}^{3\times 3}$ is the Fundamental matrix for object j and $\mathbf{x}_k = (x_k, y_k, 1)^T$ are the homogeneous coordinates of the feature in view $k = 1, 2$. Therefore, $[x_1, y_1, x_2, y_2]^T$ is a root of a second order polynomial with four variables and hence belongs to the second order algebraic variety associated with this polynomial (or, equivalently, it lies on the quadratic surface defined by all the roots of the polynomial).

4.1 Two view motion segmentation as algebraic variety clustering

Since all two-view correspondences associated with a given object satisfy (1), the problem addressed in this paper is a special case of algebraic variety clustering where the goal is to segment points lying on a surface defined by the union of n_v algebraic varieties of the form $V_i \doteq \{\mathbf{x} \colon\ p_{2,i}(\mathbf{x}) = 0\}$, where $p_{2,i}(.)$ are quadratic multivariate polynomials.

The algebraic segmentation problem (and hence two view segmentation) can be solved by first estimating the polynomials that define each variety and then assigning points $\mathbf{x_i}$ to the polynomial that yields the smallest fitting error $|p_{2,i}(\mathbf{x}_j)|$. In the case of linear subspaces, this is precisely the approach used by GPCA [15, 70]. In principle, a straightforward approach to extend GPCA to second order algebraic varieties, is to simply use a polynomial lifting to lift the problem from its original space to the space defined by the Veronese map $\mathbf{v}_2(\mathbf{x})$. Under this lifting, the problem reduces to subspace clustering, where each subspace is of the form $\mathbf{v}_2(\mathbf{x})^T \mathbf{p}_k = 0$, where the vector $\mathbf{p}_k$ contains the coefficients of the polynomial $p_{2,k}(.)$. However, even for the linear case it is well known that GPCA is fragile to noise and outliers. This situation is exacerbated when extending the approach to algebraic varieties, since the noise is polynomially lifted. In addition, this lifting ignores the specific structure of each of the polynomials that define the varieties (1), potentially introducing spurious solutions. As we show in the paper, these difficulties can be circumvented by considering a "one-at-a-time" approach that combines algebraic and Christoffel function arguments.

4.2 Approximating support sets via Christoffel polynomials

Given a probability measure μ supported on $\mathbb{R}^d$, its associated moments sequence is given by

$$m_\alpha = \mathscr{E}_\mu(\mathbf{x}^\alpha) = \int_{\mathbb{R}^d} \mathbf{x}^\alpha d\mu \tag{2}$$

where $\mathbf{x} \doteq \begin{bmatrix} x_1 & x_2 & \dots & x_d \end{bmatrix}^T$, $\alpha \doteq \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_d \end{bmatrix}$ and $\mathbf{x}^\alpha$ stands for $x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$. Each sequence m can be associated with a moment matrix $\mathbf{M}_n$, with entries $\mathbf{M}_{i,j} = m_{\alpha_i+\alpha_j}$, containing moments of order up to $2n$. In the sequel, we will use the submatrix $\mathbf{L}_n$ of $\mathbf{M}_n$, containing only moments of order $2n$. For instance, for moments of order 4 in two variables, we have

$$\mathbf{L}_2 = \begin{bmatrix} m_{(4,0)} & m_{(3,1)} & m_{(2,2)} \\ m_{(3,1)} & m_{(2,2)} & m_{(1,3)} \\ m_{(2,2)} & m_{(1,3)} & m_{(0,4)} \end{bmatrix}$$

By construction $\mathbf{L}_n \succeq 0$ thus it induces a reproducing Kernel $K_n(\mathbf{x},\mathbf{y}) \doteq \mathbf{v}_n^T(\mathbf{x})\mathbf{L}_n^{-1}\mathbf{v}_n(\mathbf{y})$[1]. The non-negative function $Q_n^{-1}(\mathbf{x}) \doteq \mathbf{v}_n^T(\mathbf{x})\mathbf{L}_n^{-1}\mathbf{v}_n(\mathbf{x})$ is known as the Christoffel function associated with the Kernel $\mathbf{K}$ [71]. It is related to the measure μ that induces $\mathbf{L}_n$ through the following optimization problem over homogeneous polynomials of degree n [71,72]:

$$\begin{aligned} p_{\mathbf{y}}^*(.) &= \underset{p \in \mathscr{P}_n^d}{argmin} \int_{\mathbb{R}^p} p^2(\xi) d\mu \;\; \text{s.t. } p(\mathbf{y}) = 1 \\ Q_n^{-1}(\mathbf{y}) &= \mathscr{E}_\mu[(p_{\mathbf{y}}^*(.))^2] \end{aligned} \tag{3}$$

where $\mathbf{y}$ is an arbitrary given data point. That is, given the data point $\mathbf{y}$, $p_{\mathbf{y}}^*(.)$ is the minimum mean square value homogeneous polynomial of degree n, subject to the constraint $p_{\mathbf{y}}^*(\mathbf{y}) = 1$, and the Christoffel function evaluated at $\mathbf{y}$, $Q_n^{-1}(\mathbf{y})$, is precisely its mean square value. In this paper, with a slight abuse of notation, we will refer to $p_{\mathbf{y}}^*(.)$ as the Christoffel *polynomials*. An explicit expression for $p_{\mathbf{y}}^*(.)$ in terms of the singular vectors $\mathbf{u}_i$ and singular values σ_i of $\mathbf{L}_n$ is given by [73]:

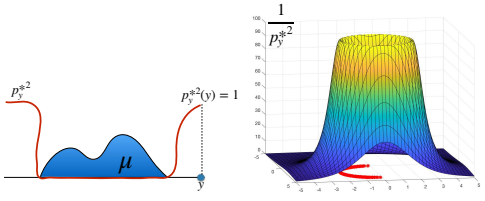
$$\begin{aligned} p_{\mathbf{y}}^*(.) &= \mathbf{v}_n(.)^T \mathbf{c}_{\mathbf{y}}^* \\ \text{where } \mathbf{c}_{\mathbf{y}}^* &= \frac{1}{\sum_{i=1}^{s_{n,d}} (\frac{1}{\sqrt{\sigma_i}} \mathbf{u}_i^T \mathbf{v}_n(\mathbf{y}))^2} \sum_{i=1}^{s_{n,d}} \frac{1}{\sigma_i} \mathbf{u}_i^T \mathbf{v}_n(\mathbf{y}) \mathbf{u}_i \end{aligned} \tag{4}$$

As noted in [72], both $Q_n(.)$ and $p_{\mathbf{y}}^*(.)$ can be used to approximate the support of the distribution μ and to detect outliers. Specifically, it can be easily shown that $\mathscr{E}_\mu(Q_n) = s_{n,d} \doteq \binom{n+d-1}{d-1}$. Direct application of Markov's inequality yields:

$$\text{prob}\{Q_n(\mathbf{y}) \geq t.s_{n,d}\} \leq \frac{1}{t} \tag{5}$$

Thus, high values of Q_n correspond to points with a high probability of being outliers. Similarly, for the polynomial $p_{\mathbf{y}}^{*2}(.)$ we have:

$$\text{prob}\left\{(p_{\mathbf{y}}^*(\mathbf{x}))^2 \geq \frac{1}{tQ_n(\mathbf{y})}\right\} \leq t \tag{6}$$



(a) (b)

Fig. 3: (a) The square Christoffel polynomial $p_{\mathbf{y}}^{*2}$ for an outlier $\mathbf{y}$ is small at inlier points. (b) The inverse of the square Christoffel polynomial, $p_{\mathbf{y}}^{*-2}$, estimated from partial data of a circle (shown in red). It has *high values at the inliers* and approximates the support of the data.

[1] For a singular $\mathbf{L} \doteq \mathbf{U} \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^T$, $\mathbf{L}^{-1} \doteq \mathbf{U} \begin{bmatrix} \Sigma^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^T$.

Remark 1 *From the equations above it follows that if* $\mathbf{y}$ *is chosen to be an outlier to the distribution* μ*, then the polynomial* $p_{\mathbf{y}}^{*2}(.)$ *will approximate the complement of the support of* μ*, in the sense that its value will be large in places where* μ *is small and viceversa (Fig. 3). This follows from the observation that if* $\mathbf{y}$ *is an outlier to the distribution* μ*, then* $Q_n(\mathbf{y})$ *is large and, from (6),* $(p_{\mathbf{y}}^*(\mathbf{x}))^2$ *is small if* $\mathbf{x}$ *is an inlier. Intuitively, if* $\mathbf{y}$ *is an outlier, a solution to (3) will be a polynomial that is close to one in a neighborhood of* $\mathbf{y}$*, to satisfy the constraint* $p(\mathbf{y}) = 1$*, and small in regions where* μ *is large, to minimize the overall integral (Fig. 3(a)). Since the region around* $\mathbf{y}$ *has low density, it contributes little to the integral of* p^2*, while setting* p^2 *small in regions where* μ *is large minimizes their cost. This observation will be key in developing the clustering algorithm.*

Note in passing that since the distribution μ is typically unknown, $\mathbf{L}_n$ cannot be computed. Rather, it is approximated by the empirical moments matrix:

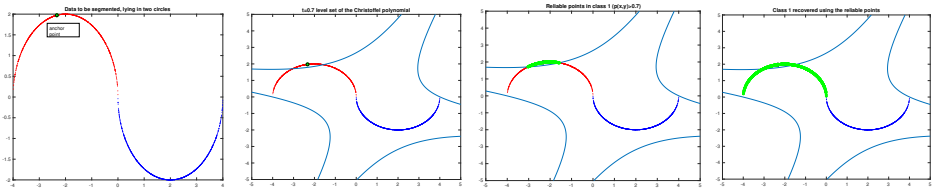
$$\mathbf{L}_n \approx \frac{1}{N}\sum_{i=1}^{N} \mathbf{v}_n(\mathbf{x})\mathbf{v}_n^T(\mathbf{x}).$$

5 Methodology and Algorithm

In this section we present a computationally efficient algorithm to segment the given data. For simplicity, we will cover the basic ideas of the algorithm for the generic algebraic variety clustering case and then indicate refinements to improve performance in the specific case of two view motion segmentation.

5.1 One at a time algebraic clustering

The proposed iterative algorithm is based on the observation made in Remark 1 that the polynomial $p_{\mathbf{y}}^*$ constructed based on a point $\mathbf{y}$ that is an outlier to the distribution μ, provides a good approximation to the support of its complement. Consider again the arrangement of n_v varieties $\mathscr{A} \doteq \cup_{j=1}^{n_v} V_j$, and suppose that we select a point $\mathbf{y}_i \in V_i$. Since this point is an outlier to the distribution over the partial arrangement $\mathscr{A}_{\text{partial}} = \cup_{j\neq i} V_j$, then it is expected that $p_{\mathbf{y}_i}^*$ will approximate, at least around $\mathbf{y}_i$ the support of its complement $\overline{\mathscr{A}}_{\text{partial}} = V_i$. Thus, a "reliable" subset $V_{i,rel}$ of V_i can be found by simply collecting points where $p_{\mathbf{y}_i}^*$ is above a threshold, related to the probability of misclassification. These "reliable" inliers can be used to refine the estimates of the coefficients of the polynomial $p_{\mathbf{y}_i}^*$ and grow the set $V_{i,rel}$ by adding new points where $p_{\mathbf{y}_i}^*$ is above the threshold. Points on this set are removed from the population and the process is repeated for the remaining varieties. A heuristic for choosing $\mathbf{y}_i$ is to select, at each stage, the point corresponding to the minimum Q. The rationale behind this choice is to select a point located in a "high mass" region of the distribution, and hence likely to have a large number of points from the same variety in its neighborhood, maximizing the number of reliable inliers used to estimate the set $V_{i,rel}$. An illustration of how to apply these ideas to a simple case is shown in Fig. 4. The corresponding conceptual Algorithm is outlined in Algorithm 1.



(a) (b) (c) (d)

Fig. 4: Applying Algorithm 1 to segment two circles. (a): Original data (red and blue points) and selected anchor (green) point (Alg. 1, step 2). (b): t=0.7 level set of the Christoffel polynomial. The anchor point and its neighborhood are outside the set (Alg. 1, step 3). (c): Points outside the level set are "reliable points", used to estimate the parameters of the circle (Alg. 1, step 4). (d): Final segmentation using the circle estimated from the reliable points (Alg. 1, step 5).

Algorithm 1 Conceptual one-at-a-time segmentation.

1: **for** $k := 1$ to $n_v - 1$ **do**
2: **Selecting** $\mathbf{y}_k$: Compute $Q_{n_v+1-k}(\mathbf{x})$ and choose $\mathbf{y}_k = \arg\min_{\mathbf{x}} Q_{n_v+1-k}(\mathbf{x})$
3: Compute $p^*_{\mathbf{y}_k}(\mathbf{x})$, the Christoffel support polynomial, for the union of $n_v - k$ clusters treating the point $\mathbf{y}_k$ as an outlier.
4: Assign points where $(p^*_{\mathbf{y}_k}(\mathbf{x}))^2 \geq t$ to the set $\mathscr{X}_k$. This set approximates a subset of V_k, the variety that contains $\mathbf{y}_k$.
5: Using the points in $\mathscr{X}_k$ and available structural information, estimate the polynomial p_k that characterizes V_k. Estimate $\hat{V}_k = \mathscr{X}_k \cup \{\mathbf{y} \colon |p_k(\mathbf{y})| \leq \text{noise level}\}$.
6: Remove the points in $\hat{V}_k$ from the population $\mathbf{x}$.

5.2 Refinements for two view motion segmentation

For the specific case of interest to this paper, the basic conceptual algorithm can be refined to improve robustness as follows. Firstly, note that the polynomial $\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2$ does not contain pure quadratic terms. Thus, when computing the moment matrix $\mathbf{L}$, rather that using the standard veronese map, one can use the restricted one:

$$\mathbf{v}_r(x_1, y_1, x_2, y_2, 1) \doteq [1 \; x_1 \; y_1 \; x_2 \; y_2 \; x_1x_2 x_1y_2 \; y_1x_2 \; y_1y_2]^T$$

This avoids spurious solutions that do not correspond to fundamental matrices when computing $p_{\mathbf{x}_{o,k}}$. Secondly, once the initial set of reliable points has been computed, it can be used to estimate the Fundamental matrix, using for instance the 8 points algorithm, possibly combined with RANSAC. In turn, this estimated Fundamental matrix can be used to find additional correspondences belonging to this object. The complete two view segmentation algorithm is given in Algorithm 2.

6 Experimental Evaluation

In this section we evaluate the performance of the proposed two-view motion segmentation method using the following benchmark datasets: Adelaide-F [74], two-view versions of Hopkins [24], with and without outliers, two-view version of KT3D dataset

Algorithm 2 Two-view motion segmentation algorithm. Lines 2-8 perform outlier rejection; 9-26 implement one at a time clustering on the estimated inliers; 27-34 assign all inlier correspondences to clusters; 35-39 use the model of each cluster to classify the unreliable points found in Step 1. Code is available at https://github.com/BengisuOzbay/TwoViewMotSeg

1: **Input:** Data matrix $\mathbf{X} \in \mathbb{R}^{5\times N}$ where each column is $[x_1, y_1, x_2, y_2, 1]^T$, number of objects M
2: **Find a reliable set for arrangement** $\mathscr{A}_M = \bigcup_{i=1}^{m} V_i$**:**
3: $t = 0.6 \times mean(Q_M)$ ▷ Initialize threshold
4: $\mathbf{X}_{rel} \leftarrow \mathbf{X}(:, Q_M < t)$ ▷ Pick the most reliable data
5: **Outlier rejection:**
6: $\mathbf{y}_o = \arg\min_{\mathbf{x}} Q_M(\mathbf{x})$ ▷ Grossest outlier
7: $t = 0.001 \times M$ ▷ Initialize threshold
8: $\mathbf{X}_{in} \leftarrow \mathbf{X}(:, p_{\mathbf{y}_o,M}(\mathbf{x} \mid \mathbf{X}_{rel}) < t)$ ▷ Remove outliers
9: **Clustering:**
10: $\mathbf{X}_{av} \leftarrow \mathbf{X}_{in}$ ▷ Initialize available
11: **for** $k := 1$ to M **do**
12: $m \leftarrow M - k$ ▷ Available clusters
13: **if** $k \neq M$ or $|\mathbf{X}_{av}| \geq 18^{\dagger}$ **then**
14: $\mathbf{y}_a \leftarrow \arg\min_{\mathbf{x}}(Q_{m+1}(\mathbf{x}))$ ▷ "anchor" point
15: $t \leftarrow$ otsu threshold for $p^2_{\mathbf{y}_a,m}(\mathbf{x})$
16: $V_{o,k} \leftarrow \mathbf{X}_{av}(:, p^2_{\mathbf{y}_a,m}(\mathbf{x}) \geq t)$ ▷ $\mathbf{y}_a \in V_{o,k} \subset V_k$
17: **else** ▷ last cluster or $|\mathbf{X}_{av}|$ is too small
18: $V_{o,k} \leftarrow \mathbf{X}_{av}$
19: **Grow the variety** $V_{o,k}$ **using Fundamental Matrix**
20: $V_{o,k,cl} \leftarrow V_{o,k}(Q_1(x \mid V_{o,k}) < mean(Q_1) \times 1.2)$ ▷ Clean the variety $V_{o,k}$
21: $\mathbf{F}_k \leftarrow norm8Point(V_{o,k,cl})$ ▷ Normalized 8 pt. alg.
22: $\mathbf{fit} \leftarrow abs(p_1^T \mathbf{F}_k p_2)$ ▷ $[p_1^T \ p_2^T \ 1]^T \in \mathbf{X}_{av}$
23: $t_F \leftarrow 0.05$, $V_{id,k} \leftarrow \mathbf{X}_{av}(:, \mathbf{fit} < t_F)$
24: **Find a reliable set** $V_{rel,k} \subset V_{id,k} \subset V_k$
25: $V_{rel,k} \leftarrow V_{id,k}(Q_1(x \mid V_{id,k}) < mean(Q_1) \times 1.4)$
26: Update available data: $\mathbf{X}_{av} \leftarrow \mathbf{X}_{av} \setminus V_{id,k}$
27: **Assign all correspondences to found clusters:**
28: **for** $j := 1$ to M **do** ▷ Find the score for each variety
29: $\mathbf{score}_Q(j) \leftarrow Q_1(\mathbf{x} \mid V_{rel,j}) \ / norm(Q_1(\mathbf{x} \mid V_{rel,j}))$
30: $\mathbf{F}_j \leftarrow norm8Point(V_{rel,j})$
31: $\mathbf{fits}_j \leftarrow abs(p_1^T \mathbf{F}_k p_2)$ ▷ $[p_1^T \ p_2^T \ 1]^T \in \mathbf{X}$
32: $V_{rel,all} = V_{rel,1} \bigcup V_{rel,2} \bigcup \ldots V_{rel,M}$
33: $\mathbf{score}_Q(0) \leftarrow (Q_M(\mathbf{x} \mid V_{rel,all}) \ / norm(Q_M(\mathbf{x} \mid V_{rel,all}))$
34: **labels** $\leftarrow \arg\min_i score_Q(j)$ ▷ Assign each pair to the cluster with lowest score, cluster $j = 0$ represents outliers
35: **Refine outliers by cross checking using** F_k
36: $[\mathbf{fits}_{min} \ \mathbf{fits}_{idx}] = min(\mathbf{fits})$ ▷ Find the cluster associated with min fit for each point
37: $\mathbf{idx}_{in} \leftarrow \mathbf{fits}_{idx}(\mathbf{fits}_{min} < 0.02)$ ▷ Inlier index according to their fit with subspaces
38: **if labels**$(\mathbf{idx}_{in}) == 0$ **then**
39: **labels**$(\mathbf{idx}_{in}) = \mathbf{idx}_{in}$

† The dimension of the modified Veronese mapping for a single cluster is 9

[28], two image pairs, named as BC and BCD, introduced in [75] and pairwise indoor scenes dataset introduced in [30]. For all datasets, we use the same thresholds as shown in lines 3 and 7 in Algorithm 2. We report the performance in terms of segmentation error [53] and computation time, and compare against the state-of-art algorithms: T-linkage [52], RCM [53], RPA [54], CBS [57], DGSAC [76], P-x [68], P-xp [69] and MLink [66]. In all cases we used code provided by their authors (except for DGSAC for which we report values from [76].

A summary of our experiments is given in Table 1, followed by a detailed analysis of each data set. As noted before, the proposed method is at least one order of magnitude faster than the best method (with comparable or better accuracy), with the ratio growing up to three orders of magnitude for larger number of correspondences. Further, it is always the fastest, with an error rate at least 50% smaller than the runner up, with this number growing up to 3 times as the number of correspondences increases.

	Adelaide-F		H-C		H-O		KT3D		BC		BCD		Pairwise	
	SE%	time	SE%	time	SE%	time	SE%	time	SE%	time	SE%	time	SE%	time
T-L	25.1	4.00	33.6	6.27	26.2	37.2	28.5	3.90	15.8	137	30.8	172	40.1	95.4
RCM	7.65	1.96	18.2	4.72	9.83	4.36	24.7	2.00	44.3	7.88	30.6	10.3	12.1	6.93
RPA	5.49	9.65	***6.41***	11.3	**8.5**	50.9	46.4	12.4	***7.56***	153	***5.3***	210	***2.63***	142
CBS	**5.03**	0.69	10.2	1.6	15.7	1.5	44.9	0.58	39.6	3.83	28.8	2.84	22.8	3.08
DGSAC*	6.95	1.27	-	-	-	-	-	-	-	-	-	-	-	-
P-x	11.5	0.30	17.6	0.39	9.77	0.71	***15.0***	0.257	10.3	1.73	27.3	1.07	18.0	0.996
P-xp	22.7	**0.06**	32.2	**0.03**	29.2	**0.29**	25.6	**0.017**	30.1	**0.08**	16.7	**0.27**	30.7	**0.091**
MLink	13.1	0.29	16.9	0.28	16.1	1.92	19.6	0.178	22.1	3.40	22.5	3.77	10.7	2.45
Ours	***4.15***	***0.02***	**7.42**	***0.02***	***7.79***	***0.03***	**17.2**	***0.016***	**9.23**	***0.05***	**7.25**	***0.09***	**4.93**	***0.036***

Table 1: Summary of our experiments: Average segmentation errors in %; 5 random runs on each scene, and the average processing times per scene (in secs) for each problem: two-view motion fitting on the two view subsets of Adelaide (2nd-3rd cols), Hopkins dataset without outliers (4th-5th) and with uniformly added outliers (6th-7th), KT3D dataset (8th-9th), a single image pair with 2 motions and 1116 feature correspondences (10th-11th), a single image pair with 3 motions and 1227 feature correspondences (12th-13th), Pairwise dataset (14th-15th) a subset of the original data set containing all pairs with more than 700 correspondences. Best results are shown in italic red and second best results are blue.*Entries for DGSAC from [76]. In all cases the proposed method is the fastest one and has the best or second best average segmentation error.

6.1 Adelaide-F

Adelaide-F is part of a larger dataset, Adelaide-RMF, widely used for homography and Fundamental matrix fitting problems. Since this paper deals with two-view motion segmentation, the only relevant portion of this dataset is the subset Adelaide-F, consisting of group of images intended for Fundamental matrix estimation problems. The Adelaide-F dataset has 19 image pairs of different sizes, with correspondences between two frames manually annotated. In each image pair there are from one to four rigid motions as well as outliers.

Fig. 1(b) compares the segmentation error, averaged over five random runs versus the average computational time per image pair. As shown there, our method has the best

performance both in terms of mean error (4.15%) and runtime (0.02secs). Our method is 30 times faster than the runner-up in error (CBS) and has $1/5$ of the error and is 3 times faster than the second fastest method, P-xp.

6.2 Hopkins-clean (H-C) and Hopkins-outliers (H-O)

The original Hopkins dataset has 156 sequences with two and three moving objects, and a single sequence with 5 moving objects. This dataset is one of the most widely used benchmarks for subspace clustering and motion segmentation problems. In the original dataset each sequence consists of 30 frames long videos of moving objects where each feature point is tracked through the entire video. [38] uses this dataset for two-view motion segmentation problem by using the first and last frames of each video as image pairs. For this experiment, following [38] we initially used the first and last

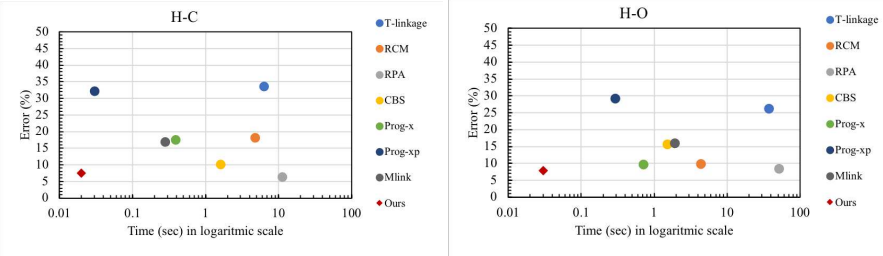


Fig. 5: Average error versus time for Hopkins (left) and Hopkins with outliers (right)

frames of the each video and referred to this as Hopkins-clean (H-C). To evaluate the effects of the displacement, we also run experiments where we selected the 1^{st} and 10^{th} frames of each sequence as the image pairs. Finally, since the Hopkins dataset does not have outliers, we generated a Hopkins-outlier (H-O) dataset, contaminated by synthetic outliers generated by declaring as matches uniformly distributed random pixels from each view. The overall results are given in Table 1 and Fig. 5, for the case where $r = (number\ of\ outliers) \div (number\ of\ inliers) = 0.9$. As shown there, our method is the second best for H-C and best for H-O, and provides the best error versus time trade-off, running more than 2 orders of magnitude faster than the method with the best/second-best error.

Robustness to outliers: In order to evaluate the performance of the proposed method under changing number of outliers, we tested it with different outlier ratios. As before, these outliers where generated by matching randomly chosen pixels from both views. The results, shown in Table 2, illustrate that the proposed method is indeed robust to outliers, since performance degrades gracefully even when the outlier ratio is substantially increased. This behavior matches the observation in [73] that algebraic methods are robust to outliers as long as enough reliable inliers are available to estimate the polynomials associated with the algebraic varieties.

	$r = 0.9$	$r = 1.5$	$r = 2.1$
SE%	7.79	9.37	10.72
time	0.03	0.03	0.04

Table 2: Hopkins dataset with varying outlier ratio where $r = (number\ of\ outliers) \div (number\ of\ inliers)$.

Displacement: To investigate the effect of displacement of the objects on the performance, we tested the proposed method and state-of-the-art methods on different

image pairs of the Hopkins-clean. Table 3 compares results using pairs of 1st-10th frames and 1st-last. As shown there, although the performance of our method slightly decreases, it is still the second best performance in terms of accuracy with the shortest run time.

Frames	1^{st}-10^{th}		1^{st}-last	
	SE%	time	SE%	time
T-L	33.4	8.25	33.6	6.27
RCM	24.8	3.87	18.2	4.72
RPA	***5.5***	14.7	***6.4***	11.4
CBS	17.9	1.5	10.2	1.6
P-x	10.7	0.28	17.6	0.39
P-xp	33.1	**0.19**	32.2	**0.03**
MLink	12.1	0.249	16.9	0.28
Ours	**9.9**	***0.014***	**7.4**	***0.02***

Table 3: Comparison of segmentation results on H-C using frames 1 and 10 versus 1 and last.

Categorical Analysis: Fig. 6 provides a detailed analysis of the Hopkins results separated by number of motions. For two motion cases our method has the second best accuracy in both clean and the outlier version of Hopkins, whereas the performance degrades in the 3-motion cases. The reasons for performance drop in our method when moving to 3 motions are: (a) In this case there are less points per object so estimation of the moment matrix is less accurate. Our method is designed to work well for scenarios with large number of correspondences, where computational complexity may render other methods impractical. (b) The results for 3-motion seem to be poisoned by the checkerboard scenes where the camera is very close to the objects and has a significant movement, while the objects have small displacements. As a consequence, the dominant motion of all objects is the same as the background category leading to clusters of two objects becoming corrupted with points belonging to the background. This is illustrated in Fig. 6 (right panel) showing that dropping this sequence leads to substantially smaller average segmentation error. In all cases our method has a lower error than P-xp, the fastest amongst existing methods. The method that yields the best error (RPA) is 500 times slower than ours.

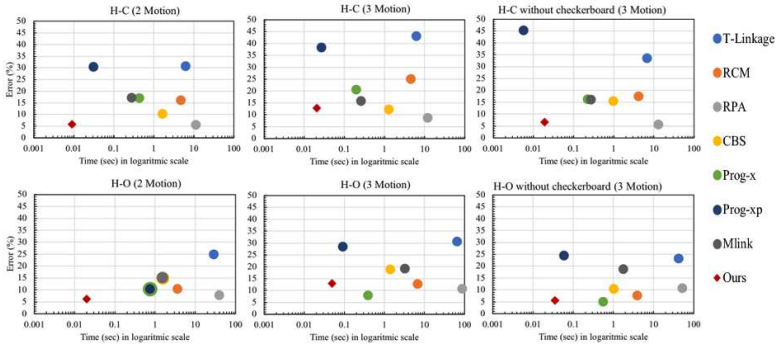


Fig. 6: Hopkins categorical analysis: H-C (top) and H-O (bottom) with 2 motion (left), 3 motion (middle) and 3 motion without checkerboard scenes (right). Segmentation error (%) and times (secs) are averaged over 5 random runs.

6.3 KT3D

KT3D was introduced by [28] as a dataset with more realistic and challenging real world effects such as strong perspective effects in the background, foreground moving objects

with limited depth reliefs, background objects with non-compact shapes, small or intermittent foreground object movement compared to that of the camera, objects moving along the epipolar line etc. We use this dataset in order to test robustness of our method against these real world challenges.The KT3D dataset has 22 videos, each 10-20 frames long with two to four moving objects. To perform 2-view motion segmentation we picked frame pairs where all the moving objects are present and frames have large perspective effects. Note that since most of the time either one or more objects appear later in the video or disappear before the last frame we could not pick the first and last frame for KT3D as we did for Hopkins. Table 1 and Fig. 7 show that our method is the fastest, with the second best error, and the best error-time trade-off. It runs one order of magnitude faster than P-x, at the price of 10% increase in error.

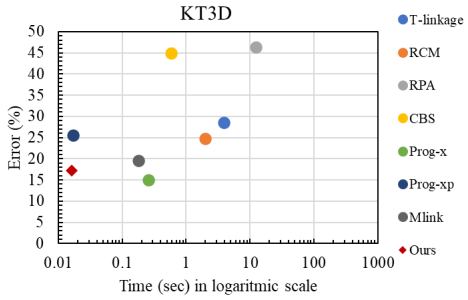


Fig. 7: Results for the KT3D data set. Our method gives the best error-time tradeoff

6.4 BC and BCD

We selected two image pairs Box-Car (BC) and Box-Car-Dinosaurs (BCD) [75] to illustrate the scaling benefits of our method. BC and BCD have two and three moving rigid objects with 1116 and 1227 feature correspondences, respectively. Both pairs are also mildly contaminated by outliers. For comparison, the Adelaide-F and Hopkins datasets have only 260 and 295 feature correspondences on average. Thus, the number of correspondences in these image pairs is significantly larger, allowing for observing its effect on the time complexity. As shown in Table 1, the proposed algorithm has the second lowest error and is the fastest method in both cases. (Fig. 8). Further, the best performing algorithm (RPA) is 3 orders of magnitude slower. Note also that RPA scales quadratically with the number of correspondences, while the proposed method scales linearly.

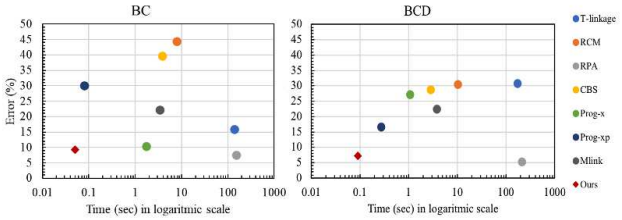


Fig. 8: Results for the KT3DMoSeg data set. Our method gives the best error-time tradeoff

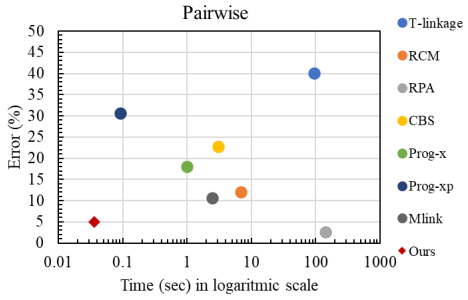


Fig. 9: Results for the Pairwise dataset. Our method gives the best error-time tradeoff

6.5 Pairwise

The Pairwise dataset was introduced in [30] as a small benchmark for pairwise matching purposes. It consists of five sequences each having six to ten different indoor scenes containing two or three motions where one object is fixed. The main challenge in these image pairs is having noisy SIFT correspondences. Since our goal is different from the original purpose for which this dataset was created [30], we used a subset consisting of 29 image pairs each having at least 700 feature correspondences to illustrate the performance and time scaling of our method. In this case the proposed method is the fastest, with second best error (Table 1 and Fig. 9). Further, the only method that has lower segmentation error is 10^3 times slower.

7 Conclusions

In this paper, we address the problem of segmenting moving rigid objects based on two-view image correspondences under a perspective camera model. We approach the problem from an algebraic perspective by exploiting the fact that, due to the geometric constraints, all correspondences from the same object lie on a quadratic surface defined by the Fundamental matrix. The proposed method is based on a characterization of each surface in terms of the Christoffel polynomial associated with the probability that a given point belongs to the surface. This allows for efficiently segmenting points "one surface at a time" by first estimating a set of points $V_{i,rel}$, the "reliable" inlier set, that, with high probability, belong to the same variety (and hence to the same object). We use these reliable inliers to estimate the parameters of the polynomial defining the variety. Finally, using this polynomial we identify additional points on the variety by simply collecting points where its absolute value is below a given threshold. It is worth emphasizing that estimating the Christoffel polynomial requires only a singular value decomposition of a matrix whose size is independent of n_p, the number of data points. Further, since this matrix is formed by adding n_p outer products, the computational complexity of the method scales linearly with the number of correspondences, which makes it well suited for applications where n_p is large.

The effectiveness of the proposed approach was demonstrated on several benchmark datasets. The experiments show that the proposed method yields the best trade-off error versus computation time. Our approach has error comparable (or better) than the most accurate method, while being at least one order of magnitude faster. In terms of execution time, the proposed algorithm is the fastest, while the runner up in execution time (P-xp) has an error rate that is at least 50% higher. Further, these gaps increase with the number of correspondences: for the largest datasets the proposed algorithm is 3 orders of magnitude faster than RPA and achieves an error less than half of the one achieved by P-xp. An additional advantage is that, since the Christoffel polynomial approximates the distribution of the inliers, it automatically provides robustness against outliers, as shown with the experiments with the Hopkins data set, where performance degraded gracefully, even in a scenario where the number of outliers was twice as high as the number of inliers.

References

1. Jong Bae Kim and Hang Joon Kim. Efficient region-based motion segmentation for a video monitoring system. *Pattern recognition letters*, 24(1-3):113–128, 2003. 1
2. Andreas Ess, Tobias Mueller, Helmut Grabner, and Luc Van Gool. Segmentation-based urban traffic scene understanding. In *BMVC*, volume 1, page 2. Citeseer, 2009. 1
3. Daniel Weinland, Remi Ronfard, and Edmond Boyer. A survey of vision-based methods for action representation, segmentation and recognition. *Computer vision and image understanding*, 115(2):224–241, 2011. 1
4. Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012. 1
5. Muhamad Risqi U Saputra, Andrew Markham, and Niki Trigoni. Visual slam and structure from motion in dynamic environments: A survey. *ACM Computing Surveys (CSUR)*, 51(2):1–36, 2018. 1
6. Pan Ji, Hongdong Li, Mathieu Salzmann, and Yuchao Dai. Robust motion segmentation with unknown correspondences. In *European conference on computer vision*, pages 204–219. Springer, 2014. 1
7. Yuxi Wang, Yue Liu, Erik Blasch, and Haibin Ling. Simultaneous trajectory association and clustering for motion segmentation. *IEEE Signal Processing Letters*, 25(1):145–149, 2017. 1
8. C William Gear. Multibody grouping from motion images. *International Journal of Computer Vision*, 29(2):133–150, 1998. 1
9. Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2790–2797. IEEE, 2009. 1
10. Ken-ichi Kanatani. Motion segmentation by subspace separation and model selection. In *Proceedings Eighth IEEE International Conference on computer Vision. ICCV 2001*, volume 2, pages 586–591. IEEE, 2001. 1
11. Jingyu Yan and Marc Pollefeys. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *European conference on computer vision*, pages 94–106. Springer, 2006. 1
12. Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):171–184, 2012. 1, 2
13. Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2765–2781, 2013. 1
14. Guangcan Liu and Shuicheng Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *2011 international conference on computer vision*, pages 1615–1622. IEEE, 2011. 1
15. R. Vidal, Y. Ma, and S. Sastry. Generalized principal component analysis (GPCA). *IEEE Trans. PAMI*, 27(12):1945–1959, 2005. 1, 5
16. Shankar Rao, Roberto Tron, Rene Vidal, and Yi Ma. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10):1832–1845, 2009. 1
17. Luca Zappella, Edoardo Provenzi, Xavier Lladó, and Joaquim Salvi. Adaptive motion segmentation algorithm based on the principal angles configuration. In *Asian Conference on Computer Vision*, pages 15–26. Springer, 2010. 1
18. Margret Keuper, Bjoern Andres, and Thomas Brox. Motion trajectory segmentation via minimum cost multicuts. In *Proceedings of the IEEE international conference on computer vision*, pages 3271–3279, 2015. 1

19. João Paulo Costeira and Takeo Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision*, 29(3):159–179, 1998. 1
20. Amit Gruber and Yair Weiss. Multibody factorization with uncertainty and missing data using the em algorithm. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004. 1
21. René Vidal, Roberto Tron, and Richard Hartley. Multiframe motion segmentation with missing data using powerfactorization and gpca. *International Journal of Computer Vision*, 79(1):85–105, 2008. 1
22. Yasuyuki Sugaya and Kenichi Kanatani. Geometric structure of degeneracy for multi-body motion segmentation. In *International Workshop on Statistical Methods in Video Processing*, pages 13–25. Springer, 2004. 1
23. Fernando Flores-Mangas and Allan D Jepson. Fast rigid motion segmentation via incrementally-complex local models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2259–2266, 2013. 1
24. Roberto Tron and René Vidal. A benchmark for the comparison of 3-d motion segmentation algorithms. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2007. 1, 8
25. Konrad Schindler, U James, and Hanzi Wang. Perspective n-view multibody structure-and-motion through model selection. In *European Conference on Computer Vision*, pages 606–619. Springer, 2006. 2
26. Ralf Dragon, Bodo Rosenhahn, and Jörn Ostermann. Multi-scale clustering of frame-to-frame correspondences for motion segmentation. In *European Conference on Computer Vision*, pages 445–458. Springer, 2012. 2
27. Zhuwen Li, Jiaming Guo, Loong-Fah Cheong, and Steven Zhiying Zhou. Perspective motion segmentation via collaborative clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1369–1376, 2013. 2
28. Xun Xu, Loong Fah Cheong, and Zhuwen Li. Motion segmentation by exploiting complementary geometric models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2859–2867, 2018. 2, 8, 12
29. Federica Arrigoni and Tomas Pajdla. Motion segmentation via synchronization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019. 2
30. Federica Arrigoni and Tomas Pajdla. Robust motion segmentation from pairwise matches. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 671–681, 2019. 2, 10, 14
31. Federica Arrigoni, Luca Magri, and Tomas Pajdla. On the usage of the trifocal tensor in motion segmentation. *Computer VisionECCV*, 2020. 2
32. Federica Arrigoni, Elisa Ricci, and Tomas Pajdla. Multi-frame motion segmentation by combining two-frame results. *International Journal of Computer Vision*, pages 1–33, 2022. 2
33. Philip HS Torr. Geometric motion segmentation and model selection. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 356(1740):1321–1340, 1998. 2
34. René Vidal, Stefano Soatto, Yi Ma, and Shankar Sastry. Segmentation of dynamic scenes from the multibody fundamental matrix. In *ECCV Workshop on Vision and Modeling of Dynamic Scenes*, 2002. 2
35. Hongdong Li. Two-view motion segmentation from linear programming relaxation. In *2007 IEEE conference on computer vision and pattern recognition*, pages 1–8. IEEE, 2007. 2

36. Yong-Dian Jian and Chu-Song Chen. Two-view motion segmentation with model selection and outlier removal by ransac-enhanced dirichlet process mixture models. *International Journal of Computer Vision*, 88(3):489–501, 2010. 2
37. Heechul Jung, Jeongwoo Ju, and Junmo Kim. Rigid motion segmentation using randomized voting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1210–1217, 2014. 2
38. Bryan Poling and Gilad Lerman. A new approach to two-view motion segmentation using global dimension minimization. *International Journal of Computer Vision*, 108(3):165–185, 2014. 2, 11
39. Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized ransac. In *Joint Pattern Recognition Symposium*, pages 236–243. Springer, 2003. 2, 3
40. Yasushi Kanazawa and Hiroshi Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *BMVC*, pages 1–10. Citeseer, 2004. 2, 3
41. Ben J Tordoff and David William Murray. Guided-mlesac: Faster image transform estimation by using matching priors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10):1523–1535, 2005. 2
42. Ondrej Chum and Jiri Matas. Matching with prosac-progressive sample consensus. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 220–226. IEEE, 2005. 2, 3
43. Aveek Shankar Brahmachari and Sudeep Sarkar. Blogs: Balanced local and global search for non-degenerate two view epipolar geometry. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1685–1692. IEEE, 2009. 2
44. Paul McIlroy, Edward Rosten, Simon Taylor, and Tom Drummond. Deterministic sample consensus with multiple match hypotheses. In *BMVC*, pages 1–11. Citeseer, 2010. 2
45. Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 2, 3
46. Etienne Vincent and Robert Laganiére. Detecting planar homographies in an image pair. In *ISPA 2001. Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis. In conjunction with 23rd International Conference on Information Technology Interfaces (IEEE Cat.*, pages 182–187. IEEE, 2001. 2
47. Marco Zuliani, Charles S Kenney, and BS Manjunath. The multiransac algorithm and its application to detect planar homographies. In *IEEE International Conference on Image Processing 2005*, volume 3, pages III–153. IEEE, 2005. 2
48. Luca Magri and Andrea Fusiello. Multiple model fitting as a set coverage problem. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3318–3326, 2016. 2
49. Wei Zhang and Jana Ksecká. Nonparametric estimation of multiple structures with outliers. In *Dynamical Vision*, pages 60–74. Springer, 2006. 2
50. Roberto Toldo and Andrea Fusiello. Robust multiple structures estimation with j-linkage. In *European conference on computer vision*, pages 537–547. Springer, 2008. 2
51. Tat-Jun Chin, David Suter, and Hanzi Wang. Multi-structure model selection via kernel optimisation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3586–3593. IEEE, 2010. 2
52. Luca Magri and Andrea Fusiello. T-linkage: A continuous relaxation of j-linkage for multi-model fitting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3954–3961, 2014. 2, 4, 10
53. Trung T Pham, Tat-Jun Chin, Jin Yu, and David Suter. The random cluster model for robust geometric fitting. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1658–1671, 2014. 2, 4, 10

54. Luca Magri and Andrea Fusiello. Robust multiple model fitting with preference analysis and low-rank approximation. In *BMVC*, volume 20, page 12, 2015. 2, 4, 10
55. Philip Hilaire Torr, Slawomir J Nasuto, and John Mark Bishop. Napsac: High noise, high dimensional robust estimation-its in the bag. In *British Machine Vision Conference (BMVC)*, 2002. 3
56. Tat-Jun Chin, Jin Yu, and David Suter. Accelerated hypothesis generation for multistructure data via preference analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):625–638, 2011. 3
57. Ruwan Tennakoon, Alireza Sadri, Reza Hoseinnezhad, and Alireza Bab-Hadiashar. Effective sampling: Fast segmentation using robust geometric model fitting. *IEEE Transactions on Image Processing*, 27(9):4182–4194, 2018. 3, 10
58. Rahul Raguram, Jan-Michael Frahm, and Marc Pollefeys. A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. In *European Conference on Computer Vision*, pages 500–513. Springer, 2008. 4
59. T Kim and W Yu. Performance evaluation of ransac family. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 1–12, 2009. 4
60. Florian Kluger, Eric Brachmann, Hanno Ackermann, Carsten Rother, Michael Ying Yang, and Bodo Rosenhahn. Consac: Robust multi-model fitting by conditional sample consensus. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4634–4643, 2020. 4
61. Maksym Ivashechkin, Daniel Barath, and Jiří Matas. Vsac: Efficient and accurate estimator for h and f. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15243–15252, 2021. 4
62. Ruwan Tennakoon, David Suter, Erchuan Zhang, Tat-Jun Chin, and Alireza Bab-Hadiashar. Consensus maximisation using influences of monotone boolean functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2866–2875, 2021. 4
63. Luca Magri and Andrea Fusiello. Fitting multiple heterogeneous models by multi-class cascaded t-linkage. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7460–7468, 2019. 4
64. Guobao Xiao, Hanzi Wang, Taotao Lai, and David Suter. Hypergraph modelling for geometric model fitting. *Pattern Recognition*, 60:748–760, 2016. 4
65. Shuyuan Lin, Guobao Xiao, Yan Yan, David Suter, and Hanzi Wang. Hypergraph optimization for multi-structural geometric model fitting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8730–8737, 2019. 4
66. Luca Magri, Filippo Leveni, and Giacomo Boracchi. Multilink: Multi-class structure recovery via agglomerative clustering and model selection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1853–1862, 2021. 4, 10
67. Hanzi Wang, Guobao Xiao, Yan Yan, and David Suter. Searching for representative modes on hypergraphs for robust geometric model fitting. *IEEE transactions on pattern analysis and machine intelligence*, 41(3):697–711, 2018. 4
68. Daniel Barath and Jiri Matas. Progressive-x: Efficient, anytime, multi-model fitting algorithm. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3780–3788, 2019. 4, 10
69. Daniel Barath, Denys Rozumny, Ivan Eichhardt, Levente Hajder, and Jiri Matas. Progressive-x+: Clustering in the consensus space. *arXiv preprint arXiv:2103.13875*, 2021. 4, 10
70. R. Vidal, S. Soatto, Y. Ma, and S. Sastry. An algebraic geometric approach to the identification of a class of linear hybrid systems. In *42nd Conf. on Dec. and Control*, volume 1, pages 167–172, 2003. 5

71. Y. Xu. On orthogonal polynomials in several variables. *Special functions, q-series and related topics, The Fields Institute for Research in Mathematical Sciences, Communications Series*, 14:247–270, 1997. 6
72. Edouard Pauwels and Jean B Lasserre. Sorting out typicality with the inverse moment matrix sos polynomial. In *Advances in Neural Information Processing Systems*, pages 190–198, 2016. 6
73. M. Sznaier and O. Camps. Sos-rsc: A sum-of-squares polynomial approach to robustifying subspace clustering algorithms. In *IEEE CVPR*, pages 8033–8041, 2018. 6, 11
74. Hoi Sim Wong, Tat-Jun Chin, Jin Yu, and David Suter. Dynamic and hierarchical multi-structure geometric model fitting. In *2011 International Conference on Computer Vision*, pages 1044–1051. IEEE, 2011. 8
75. Hanzi Wang, Tat-Jun Chin, and David Suter. Simultaneously fitting and segmenting multiple-structure data with outliers. *IEEE transactions on pattern analysis and machine intelligence*, 34(6):1177–1192, 2011. 10, 13
76. Lokender Tiwari and Saket Anand. Dgsac: Density guided sampling and consensus. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 974–982. IEEE, 2018. 10