

Robust Algebraic Segmentation of Mixed Rigid-Body and Planar Motions from Two Views

Shankar R. Rao · Allen Y. Yang · S. Shankar Sastry · Yi Ma

Received: 4 September 2008 / Accepted: 9 December 2009 / Published online: 7 January 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract This paper studies segmentation of multiple rigid-body motions in a 3-D dynamic scene under perspective camera projection. We consider dynamic scenes that contain both 3-D rigid-body structures and 2-D planar structures. Based on the well-known epipolar and homography constraints between two views, we propose a hybrid perspective constraint (HPC) to unify the representation of rigid-body and planar motions. Given a mixture of K hybrid perspective constraints, we propose an algebraic process to partition image correspondences to the individual 3-D motions, called Robust Algebraic Segmentation (RAS). Particularly, we prove that the joint distribution of image correspondences is uniquely determined by a set of $(2K)$ -th degree polynomials, a global signature for the union of K motions of possibly mixed type. The first and second deriva-

tives of these polynomials provide a means to recover the association of the individual image samples to their respective motions. Finally, using robust statistics, we show that the polynomials can be robustly estimated in the presence of moderate image noise and outliers. We conduct extensive simulations and real experiments to validate the performance of the new algorithm. The results demonstrate that RAS achieves notably higher accuracy than most existing robust motion-segmentation methods, including random sample consensus (RANSAC) and its variations. The implementation of the algorithm is also two to three times faster than the existing methods. The implementation of the algorithm and the benchmark scripts are available at <http://perception.csl.illinois.edu/ras/>.

Keywords Motion segmentation · Epipolar geometry · Homography · Outlier rejection · Influence function · Algebraic segmentation

S.R. Rao (✉) · Y. Ma
Department of ECE, University of Illinois at Urbana-Champaign,
Coordinate Science Laboratory, Urbana, IL 61801, USA
e-mail: srrao@illinois.edu

Y. Ma
e-mail: yima@illinois.edu

A.Y. Yang · S.S. Sastry
Department of EECS, University of California, Cory Hall,
Berkeley, CA 94720, USA

A.Y. Yang
e-mail: yang@eecs.berkeley.edu

S.S. Sastry
e-mail: sastry@eecs.berkeley.edu

S.R. Rao
HRL Laboratories, LLC, Malibu, CA 90265, USA

Y. Ma
Visual Computing Group, Microsoft Research Asia, Beijing,
China

1 Introduction

Structure from Motion (SFM) is one of the fundamental problems in computer vision. Given a 3-D dynamic scene with multiple moving objects, the task in SFM is to recover the parameters of the motions from a sequence of 2-D images. In order to recover the motion parameters, one must assume a certain motion model for individual objects in the scene. Over the years, many different motion models have been studied, which range from assumptions on the camera projection model (e.g., perspective, affine, orthographic) to assumptions on the objects in the scene (e.g., planar or general rigid bodies). In the literature, previous investigation had been largely focused on segmentation and estimation of specific motion models, and the methods did not generalize

well to complex dynamic scenes containing multiple motions or multiple types of motions.

Particularly, the presence of planar structures complicates the estimation problem. In SFM, when a rigid object is observed in two perspective camera views, its feature correspondences satisfy either an *epipolar* constraint for a general rigid-body or a *homography* constraint for a planar object. Clearly, a planar motion is a degenerate rigid-body motion. Since planar structures commonly exist in urban environments (e.g., ground, building facade, and background in distance) together with general rigid-body objects (e.g., cars), in a typical image sequence, one should expect to see more than one type of motion, described by multiple epipolar constraints and homographies.

As an attempt to unify the analyses of different motion models, a *joint image space* has been proposed in the literature (Shapiro 1995; Triggs 1995; Anandan and Avidan 2000; Tong et al. 2004). The main focus of the study has been the representation of the epipolar constraints and/or the affine projection constraints in a joint image space. Anandan and Avidan (2000) showed that a single epipolar constraint represents a 4-D point cone in \mathbb{R}^5 , and an affine constraint represents a 3-D affine subspace in \mathbb{R}^5 . Regardless of the difference in these two ways of embedding the image pair into a joint image space, estimation of one or multiple epipolar and homography relations is converted to fitting algebraic manifolds defined by a set of implicit polynomials. The problem of fitting conics is indeed more difficult than fitting linear subspaces represented by affine projection constraints. Several inspiring solutions have been proposed to tackle this problem. The reader is referred to Goshen et al. (2005) for a review of work on estimating single epipolar constraints.

Coupled with the problem of mixture motions in a dynamic scene, the measurement of 3-D feature correspondence in the images can be noisy in terms of pixel coordinates. In addition, mismatched feature points from the objects and the background often introduce outliers in most real-world applications. The presence of image noise and outliers have historically affected the accuracy of SFM estimation. Previous studies have produced several robust algorithms that can estimate single rigid-body motions in the presence of large number of outliers. However, the extensions of these algorithms to multiple rigid-body motions do not generally perform well, because the single fundamental matrix assumption is violated in the dynamic scene (Tong et al. 2004; Schindler and Suter 2005; Yang et al. 2006).

In this paper, we study the segmentation problem of multiple rigid-body and planar motions from two views under perspective projection, one of the most general assumptions in SFM.¹ Assuming that the number of mixed motions

(K) is given, we propose a unified *hybrid perspective constraint* (HPC) in the joint image space to characterize mixture epipolar and homography constraints. The new algorithm, called *robust algebraic segmentation* (RAS), focuses on recovering a set of polynomials vanishing on the algebraic set of multiple epipolar and homography constraints. We prove, for the first time, that the distribution of K mixture epipolar and homography constraints in the joint image space is uniquely determined by a set of $(2K)$ -th degree vanishing polynomials. Based on robust statistics, we further show that these polynomials can be robustly estimated from a data set with moderate data noise and outliers. Using the polynomials, we provide a solution to segment the inlying image features into corresponding motions. Finally, a comparison is conducted to quantitatively measure the performance of RAS with several established algorithms for motion segmentation (Torr and Zisserman 2000; Schindler and Suter 2005; Subbarao and Meer 2006; Rao et al. 2008). The implementation of the algorithm and the benchmark scripts are available at <http://perception.csl.illinois.edu/ras/>.

1.1 Relations to Previous Work

To put our contributions in context, we briefly review existing methods developed to model data points sampled from a mixture of geometric manifolds or statistical distributions. An extension to *principal component analysis* (PCA) called probabilistic PCA fits data points into a mixture of multiple subspace models, using a mixture Gaussian distribution and *expectation maximization* (EM) (Tipping and Bishop 1999). Agrawal et al. (1998) presented a scalable algorithm to perform clustering on multiple individual subspaces. K -subspaces, proposed by Lee et al. (2005), is an extension of the K -means for fitting multiple linear models. Leonardis et al. (2002) developed a subspace selection method for multiple subspaces. Scholkopf et al. (1998) introduced a set of nonlinear mappings called *kernels* that map the original nonlinear data points into another space in which the manifold is linear. Recently, an agglomerative method based on the principles of lossy data compression, called *agglomerative lossy compression* (ALC) (Ma et al. 2007), has proven highly effective in segmenting data from multiple subspaces. The method has been applied to partition affine camera motions with outlying and corrupted trajectories (Rao et al. 2008). There has also been work to apply manifold learning techniques such as ISOMAP and *local linear embedding* to cluster data drawn from multiple manifolds (Souvenir and Pless 2007; Goh and Vidal 2007). In the literature of computer vision, many effective methods have been developed based on the above clustering techniques

¹We note that there are many proposed methods in the literature for performing motion segmentation under perspective projection with three

or more views (Ozden et al. 2007; Vidal and Hartley 2008). We do not compare with such methods in this paper.

to segment specific types of rigid-body motions from images (Costeira and Kanade 1998; Han and Kanade 2000; Shashua and Levin 2001; Kanatani 2002b; Kanatani and Sugaya 2003; Yan and Pollefeys 2006).²

The methods proposed in this paper have strong connections to the algebraic method pioneered by Vidal et al. (2005) known as *generalized principal component analysis* (GPCA). GPCA is a non-iterative method for segmenting data drawn from multiple linear subspaces that first estimates a set of polynomials that vanish on the data set, and then uses derivatives of these vanishing polynomials to segment samples to their respective subspaces. GPCA has been applied to motion segmentation under the affine camera model, and has also been applied to motion segmentation under the perspective camera model via various nonlinear and complex embeddings (Vidal et al. 2006; Vidal and Ma 2006; Yang et al. 2005). However, GPCA faces two major problems when applied to perspective motion segmentation. First, the embeddings used by GPCA for perspective motion segmentation assume that all the motions in the scene are of the same type (i.e., all planar motions or all general rigid-body motions), and so have no guarantees to succeed in the mixture motion case described in this paper. Second, these embeddings are also somewhat brittle, often failing to segment motion data even with little or no noise present. Our prior work has quantitatively demonstrated the inefficacy of GPCA for segmentation of mixture motions on both simulated and real scenes (Yang et al. 2005; Rao et al. 2005).

When the data are further contaminated by outliers, robust statistical methods should be applied to separate the outliers from the inliers. Outlying measurements affect model estimation and 3-D reconstruction in most computer vision applications. Despite centuries of study in robust statistics, there is no universally accepted definition for “outliers.” Most robust solutions are based on one of the following three assumptions (Yang et al. 2006):

1. Probability-based: Outliers are a set of *small-probability* samples with respect to the probability distribution in question. The given data set is therefore an atypical set if such samples constitute a significant portion of the data. Methods in this approach include the *M-estimator* (Campbell 1980; Huber 1981) and its variation, *multivariate trimming* (MVT) (Gnanadesikan and Kettenring 1972).

²Notice that there exist statistical algorithms (e.g., Anandan and Avidan 2000) that can be modified to iteratively estimate multiple quadratic manifolds. We did not compare our algorithm with this line of development. Instead, we point out that our algorithm gives a good noniterative approximation of the globally optimal solution, and it can be used to initialize other statistical algorithms so that the nonlinear optimization can be conducted near its global optimum.

2. Influence-based: Outliers are samples that have relatively *large influence* on the estimated model parameters (Campbell 1978; Hampel et al. 1986; Jolliffe 2002). The influence of a sample is normally defined as the difference between the model estimated with and without the sample.
3. Consensus-based: Outliers are samples that are *not consistent* with the model inferred from the remainder of the data. A measure of inconsistency is normally the error residue of the sample in question with respect to the model. Methods in this approach include Hough transform (Ballard 1981), least median estimators (Rousseeuw 1984), and random sample consensus (RANSAC) (Fischler and Bolles 1981; Stewart 1995; Torr and Davidson 2003; Rousseeuw 1984; Wang and Suter 2004; Schindler and Suter 2005; Frahm and Pollefeys 2006).

In robust statistics, the *breakdown point*, the minimal percentage of outliers in a data set that can cause arbitrarily large estimation error, is an important index of a robust estimator. It can be shown that the theoretical breakdown points for probability-based and influence-based methods are 50% (Hampel et al. 1986; Stewart 1999). This drawback motivates the investigation of consensus-based methods. These techniques treat outliers as samples drawn from a model that is very different from the model of inliers. Therefore, although the outlier percentage may be greater than 50%, they may not result in a model with higher consensus than the inlier model.

In computer vision, consensus-based algorithms have been popularly used to reject outliers in SFM applications, most of which are largely based on the RANSAC principle. Particularly, a series of work has been proposed to extend RANSAC from single motions (Torr 1998; Torr and Zisserman 2000; Torr and Davidson 2003) to mixture rigid-body and homography motions (Bartoli 2001; Tong et al. 2004; Schindler and Suter 2005; Subbarao and Meer 2006). On the other hand, due to the lack of a global representation for mixture motions, most algorithms have to iteratively estimate individual motions from the data set when a mixture of multiple motions is present in a common scene.

1.2 Organization of the Paper

The paper is organized as follows: Sect. 2 derives a unified quadratic polynomial constraint for both rigid-body and planar motions, and further proposes a *hybrid perspective constraint* for cases where multiple mixture motions are present. Section 3 discusses robust estimation of a single well-chosen vanishing polynomial, called the *segmentation polynomial*, from a data set with both noise and outliers. Combining these results, Sect. 4 proposes a *robust algebraic segmentation* (RAS) algorithm to segment image points of mixture

motions using the recovered segmentation polynomial. Section 5 shows that RAS complements the traditional random sampling-based methods. In particular, we apply RANSAC as a post-processing step to RAS to further improve the robustness of RAS to outliers. By using RAS in conjunction with RANSAC, we obtain much improved speed and accuracy compared to using RANSAC alone. Finally, Sects. 6 and 7 conduct extensive experiments on synthesized data and real images to evaluate the performance of the algorithm, with careful comparison to other existing methods.

2 Mixture Motions and Hybrid Perspective Constraints

2.1 Two-View Geometry on Joint Image Spaces

Given a point X in space and its image correspondences $(x_1, x_2) \in \mathbb{R}^3$ in homogeneous coordinates, its (uncalibrated) epipolar constraint with respect to a fundamental matrix

$$F \doteq \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \in \mathbb{R}^{3 \times 3}$$

is

$$x_2^T F x_1 = 0. \tag{1}$$

If X lies on a plane in space, the image correspondences also satisfy a homography constraint w.r.t. a homography matrix

$$H \doteq \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \in \mathbb{R}^{3 \times 3}$$

is

$$\hat{x}_2 H x_1 = 0 \in \mathbb{R}^3, \tag{2}$$

where $\hat{x} \in \mathbb{R}^{3 \times 3}$ is a skew-symmetric matrix of x such that $\hat{x}_2 x_1 = x_2 \times x_1$, the cross product of the two vectors. Since $\text{rank}(\hat{x}) = 2$, two of the three equations in (2) are linearly independent.

Definition 1 (Quadratic Manifold) In a D -dimensional ambient space, a *quadratic manifold*³ of dimension d is a set of points S that all jointly satisfy $D - d$ quadratic equations of the form

$$y^T Q_j y = 0, \quad \forall y \in S, \quad j = 1, \dots, D - d, \tag{3}$$

where each Q_j is a symmetric $D \times D$ matrix.

³A quadratic manifold here is an algebraic surface that satisfies a set of quadratic equations. Strictly speaking, such a surface could be an algebraic surface of order higher than two.

Constraints (1) and (2) are both bilinear in terms of the image pair (x_1, x_2) in \mathbb{R}^3 . However, we can stack $x_1 = (x_1, y_1, 1)^T$ and $x_2 = (x_2, y_2, 1)^T$ into a single vector of the form

$$y = (x_1, y_1, x_2, y_2, 1)^T \in \mathbb{R}^5. \tag{4}$$

The set $\mathcal{Y} = \{y\} \subset \mathbb{R}^5$ is a *joint image manifold* in the 5-D *joint image space* (Goshen et al. 2005).

Using the notation in (4), the epipolar equation (1) can be rewritten in the following quadratic form:

$$y^T A y \doteq y^T \begin{bmatrix} 0 & 0 & f_{11} & f_{21} & f_{31} \\ 0 & 0 & f_{12} & f_{22} & f_{32} \\ f_{11} & f_{12} & 0 & 0 & f_{13} \\ f_{21} & f_{22} & 0 & 0 & f_{23} \\ f_{31} & f_{32} & f_{13} & f_{23} & 2f_{33} \end{bmatrix} y = 0. \tag{5}$$

Equation (5) shows that joint image samples that satisfy the epipolar constraint lie on a 4-D quadratic manifold in \mathbb{R}^5 (Anandan and Avidan 2000), which is referred to as a *quadratic fundamental manifold* (QFM) in this paper.

Similarly, we can also rewrite (2) in quadratic form:

$$y^T B_1 y \doteq y^T \begin{bmatrix} 0 & 0 & 0 & h_{31} & -h_{21} \\ 0 & 0 & 0 & h_{32} & -h_{22} \\ 0 & 0 & 0 & 0 & 0 \\ h_{31} & h_{32} & 0 & 0 & h_{33} \\ -h_{21} & -h_{22} & 0 & h_{33} & -2h_{23} \end{bmatrix} y = 0,$$

$$y^T B_2 y \doteq y^T \begin{bmatrix} 0 & 0 & -h_{31} & 0 & h_{11} \\ 0 & 0 & -h_{32} & 0 & h_{12} \\ -h_{31} & -h_{32} & 0 & 0 & -h_{33} \\ 0 & 0 & 0 & 0 & 0 \\ h_{11} & h_{12} & -h_{33} & 0 & 2h_{13} \end{bmatrix} y = 0,$$

$$y^T B_3 y \doteq y^T \begin{bmatrix} 0 & 0 & h_{21} & -h_{11} & 0 \\ 0 & 0 & h_{22} & -h_{12} & 0 \\ h_{21} & h_{22} & 0 & 0 & h_{23} \\ -h_{11} & -h_{12} & 0 & 0 & -h_{13} \\ 0 & 0 & h_{23} & -h_{13} & 0 \end{bmatrix} y = 0. \tag{6}$$

In the \mathbb{R}^5 joint image space, the above three quadratic constraints are linearly independent. Hence, they define a 2-D quadratic manifold in \mathbb{R}^5 , which is referred to as a *quadratic homography manifold* (QHM). It is easy to see that any QFM or QHM shall pass the origin at $y = 0$.

To derive a single perspective constraint for either a QFM or a QHM, we define a symmetric 5-by-5 matrix

$$Q \doteq \begin{bmatrix} 0 & 0 & q_1 & q_2 & q_3 \\ 0 & 0 & q_4 & q_5 & q_6 \\ q_1 & q_4 & 0 & 0 & q_7 \\ q_2 & q_5 & 0 & 0 & q_8 \\ q_3 & q_6 & q_7 & q_8 & q_9 \end{bmatrix}, \tag{7}$$

which has *at most* 9 degrees of freedom. Then, a QFM defined in (5) is rewritten as

$$S^F \doteq \{y \in \mathbb{R}^5 : g^F(y) \doteq y^T Q^F y = 0\}. \tag{8}$$

Similarly, a QHM defined in (6) is rewritten as

$$\begin{aligned} S^H &\doteq \{y \in \mathbb{R}^5 : g_i^H(y) \doteq y^T Q_i^H y = 0, i = 1, 2, 3\}, \\ &= \{g_1^H(y) = 0\} \cap \{g_2^H(y) = 0\} \cap \{g_3^H(y) = 0\}, \end{aligned} \tag{9}$$

Subsequently, we also assume that the polynomials $g_1^H, g_2^H,$ and g_3^H are linearly independent.

The following theorem gives an important rank condition to matrix Q from the QFM or QHM constraint:

Theorem 1 *If a matrix Q satisfies either a QFM constraint (8) or a QHM constraint (9), then $\text{rank}(Q) \leq 4$.*

Proof The conclusion is straightforward given that at least one eigenvalue of Q is zero from (8) or (9). For more details, please refer to Anandan and Avidan (2000). \square

Finally, we consider a dual relation between an algebraic set S and its generating polynomials. In (8) and (9), it is clear that given a quadratic polynomial g^F , all points $y \in \mathbb{R}^5$ that *vanish* on g^F are uniquely determined, i.e., S^F . Likewise, S^H is uniquely determined by (g_1^H, g_2^H, g_3^H) . On the other hand, we are interested in a dual problem: Given an algebraic set S^F or S^H , are its generating polynomial(s) also uniquely determined?

In algebraic geometry, the quadratic polynomials (8) and (9) are said to lie in a *ring* of polynomials of five variables (Harris 1992), which is denoted by

$$R \doteq \mathbb{R}^{[5]} = \mathbb{R}[X_1, X_2, \dots, X_5]. \tag{10}$$

In this polynomial ring, all polynomials p that *vanish* on an algebraic set S constitute a special subset $I(S)$, i.e.,

$$I(S) = \{p : p(y) = 0 \text{ for all } y \in S\}. \tag{11}$$

The subset $I(S)$ is called an *ideal*. Particularly, if $g \in I(S)$, then for an arbitrary polynomial $h \in R$, clearly gh is also an element of $I(S)$ since $gh(y) = 0$ for all $y \in S$.

On the other hand, we define the generating ideals $I^F = \langle g^F \rangle$ and $I^H = \langle g_1^H, g_2^H, g_3^H \rangle$. Hence, for any polynomial $p \in I^F$,

$$p = p_1 \cdot g^F, \quad p_1 \in R. \tag{12}$$

Similarly, for any polynomial $p \in I(S^H)$,

$$p = p_1 \cdot g_1^H + p_2 \cdot g_2^H + p_3 \cdot g_3^H, \quad p_1, p_2, p_3 \in R. \tag{13}$$

The following theorem establishes the equivalence relation between $I(S)$ and its corresponding generating ideal:

Theorem 2

$$I^F = I(S^F), \quad I^H = I(S^H). \tag{14}$$

Proof See Appendix. \square

Although the proof of the theorem is technical, the theorem essentially establishes the following equivalence relation: Given a set of points $\mathcal{Y} = \{y\} \subset \mathbb{R}^5$ that satisfy a QFM or QHM constraint, its quadratic generating polynomials g_i are uniquely determined.

In summary, the set of quadratic polynomials g_i as a global signature completely determine the algebraic set S . Therefore, the problem of recovering individual QFM and QHM constraints from \mathcal{Y} is equivalent to that of recovering their quadratic vanishing polynomials. For more detailed development between an algebraic set and its vanishing polynomials, the reader is referred to Ma et al. (2008).

2.2 Hybrid Perspective Constraints

In this paper, our focus is on an image set \mathcal{Y} from a mixture of motions. Suppose $\mathcal{Y} = \{y_1, \dots, y_N\}$ is drawn from a union of QFM’s and QHM’s:

$$\mathcal{A} \doteq S_1 \cup S_2 \cup \dots \cup S_K, \tag{15}$$

where each S_i may represent a QFM or a QHM. Then the problem of motion segmentation becomes the problem of simultaneously segmenting the data set \mathcal{Y} into its individual motions sets S_i , and recovering the motion coefficients represented by fundamental matrices F ’s and homography matrices H ’s.

In Sect. 2.1, we have shown that given a single motion S , its vanishing polynomials (8) or (9) completely determine the algebraic set. Similarly, with a mixture motion \mathcal{A} , we are also looking for a set of polynomials of higher degrees that serve as a global signature of the mixture motion. The following theorem proves the existence of such polynomials:

Theorem 3 *Let I_{2K} denote the set of polynomials in $(2K)$ -th degree that vanish on a mixture motion $\mathcal{A} = S_1 \cup S_2 \cup \dots \cup S_K$.*

1. I_{2K} is a subspace (of $(2K)$ -degree polynomials).
2. \mathcal{A} is uniquely determined by I_{2K} .

Proof See Ma et al. (2008). \square

We reiterate Theorem 3 in the context of motion segmentation: In order to estimate the global structure of \mathcal{A} as a union of mixture motions, one only needs to recover a set I_{2K} of $(2K)$ -th degree polynomials that vanish on the data

set \mathcal{Y} . This is a very small subset of all the vanishing polynomials on \mathcal{A} of arbitrary degrees. Furthermore, since I_{2K} is a polynomial subspace, one only needs to recover a maximal set of linearly independent polynomials as its basis. Hence, we define a *hybrid perspective constraint* as follows:

Definition 2 (Hybrid Perspective Constraint) A *hybrid perspective constraint* (HPC) for a union of K mixture epipolar and homography motions \mathcal{Y} is a maximal set of linearly independent $(2K)$ -th degree polynomials that vanish on \mathcal{Y} .

As shown in Theorem 3, the HPC is both a *necessary* and *sufficient* global polynomial constraint that represents a union of mixture motions. In other words, \mathbf{y} is a joint image sample from a union of mixture motions \mathcal{A} if and only if all of the polynomials in the corresponding HPC for \mathcal{A} vanish on \mathbf{y} . We illustrate the HPC in the following example.

Example 1 Suppose that the underlying distribution is a mixture of one QFM, S^F , and one QHM, S^H . Then samples from the QFM satisfy a quadratic constraint: $\mathbf{y}^T Q_{1,1} \mathbf{y} = 0$; and samples from the QHM satisfy three quadratic constraints: $\{\mathbf{y}^T Q_{2,1} \mathbf{y} = 0\}$, $\{\mathbf{y}^T Q_{2,2} \mathbf{y} = 0\}$, and $\{\mathbf{y}^T Q_{2,3} \mathbf{y} = 0\}$. Therefore, all samples in the mixture must simultaneously satisfy the following three vanishing polynomials:

$$\begin{aligned} p_1(\mathbf{y}) &= (\mathbf{y}^T Q_{1,1} \mathbf{y})(\mathbf{y}^T Q_{2,1} \mathbf{y}), \\ p_2(\mathbf{y}) &= (\mathbf{y}^T Q_{1,1} \mathbf{y})(\mathbf{y}^T Q_{2,2} \mathbf{y}), \\ p_3(\mathbf{y}) &= (\mathbf{y}^T Q_{1,1} \mathbf{y})(\mathbf{y}^T Q_{2,3} \mathbf{y}). \end{aligned}$$

Hence, the quadratic polynomials p_1, p_2, p_3 of degree 4 are a basis for $I_4(S^F \cup S^H)$, and they simultaneously vanish on both S^F and S^H . Finally, because p_1, p_2, p_3 are linearly independent, $\dim(I_4(S^F \cup S^H)) = 3$.

In the rest of this section, by assuming a perfect data set \mathcal{Y} without noise and outliers, we introduce a means of estimating from \mathcal{Y} a maximal set of $(2K)$ -th degree independent polynomial as a basis of I_{2K} .

Definition 3 (Veronese Map) Given K independent motions, the *Veronese map* of degree $(2K)$, $v_{2K} : \mathbb{R}^5 \rightarrow \mathbb{R}^{M_K^5}$, is defined as

$$v_{2K} : (x_1, y_1, x_2, y_2, z)^T \mapsto (x_1^{2K}, x_1^{2K-1}y_1, \dots, z^{2K})^T, \tag{16}$$

where $M_K^5 = \binom{2K+5-1}{2K}$ is the total number of monomials of degree $(2K)$ with 5 variables, i.e., the dimension of the codomain of v_{2K} .

Using the Veronese map, we can rewrite an arbitrary homogeneous polynomial $p(\mathbf{y})$ of $(2K)$ -th degree as

$$p(\mathbf{y}) = \mathbf{c}^T v_{2K}(\mathbf{y}), \tag{17}$$

where \mathbf{c} consists of all coefficients of $p(\mathbf{y})$ associated with the monomials in $v_{2K}(\mathbf{y})$.

Veronese maps have been well studied in the segmentation of subspace arrangements, i.e., a finite union of linear subspaces (Vidal et al. 2005; Ma et al. 2008). However, we notice that a vanishing polynomial for an HPC is *not* a general $(2K)$ -th degree polynomial. Particularly, as shown in (7), each Q matrix only has 9 degrees of freedom. Referring back to (5) and (6), the two 2-by-2 diagonal minors are zero for all Q matrices. Enforcing the zero coefficients in recovering the vanishing polynomials is essential to achieve accurate estimation of the individual motions.

Example 2 Let $\mathbf{y} \in S$ be joint image samples drawn from two rigid body motions parameterized by the matrices

$$Q_1 = \begin{bmatrix} 0 & 0 & a_1 & b_1 & c_1 \\ 0 & 0 & d_1 & e_1 & f_1 \\ a_1 & d_1 & 0 & 0 & g_1 \\ b_1 & e_1 & 0 & 0 & h_1 \\ c_1 & f_1 & g_1 & h_1 & i_1 \end{bmatrix} \quad \text{and} \quad Q_2 = \begin{bmatrix} 0 & 0 & a_2 & b_2 & c_2 \\ 0 & 0 & d_2 & e_2 & f_2 \\ a_2 & d_2 & 0 & 0 & g_2 \\ b_2 & e_2 & 0 & 0 & h_2 \\ c_2 & f_2 & g_2 & h_2 & i_2 \end{bmatrix},$$

respectively. Then, each sample $\mathbf{y} = (x_1, y_1, x_2, y_2, 1)^T$ satisfies

$$\begin{aligned} p(\mathbf{y}) &= (\mathbf{y}^T Q_1 \mathbf{y})(\mathbf{y}^T Q_2 \mathbf{y}) \\ &= (2a_1x_1x_2 + 2b_1x_1y_2 + 2c_1x_1 + 2d_1y_1x_2 \\ &\quad + 2e_1y_1y_2 + 2f_1y_1 + 2g_1x_2 + 2h_1y_2 + i_1) \\ &\quad \times (2a_2x_1x_2 + 2b_2x_1y_2 + 2c_2x_1 + 2d_2y_1x_2 \\ &\quad + 2e_2y_1y_2 + 2f_2y_1 + 2g_2x_2 + 2h_2y_2 + i_2). \end{aligned}$$

Though there are 70 monomials of order at most 4, there are only 36 such monomials with nonzero coefficients in $p(\mathbf{y})$: $x_1^2x_2^2, x_1^2x_2y_2, x_1^2x_2, x_1^2y_2^2, x_1^2y_2, x_1^2, x_1y_1x_2^2, x_1y_1x_1x_2, x_1y_1x_2, x_1y_1y_2^2, x_1y_1y_2, x_1y_1, x_1x_2^2, x_1x_2y_2, x_1x_2, x_1y_2^2, x_1y_2, x_1, y_1^2x_2^2, y_1^2x_2y_2, y_1^2x_2, y_1^2y_2^2, y_1^2y_2, y_1^2, y_1x_2^2, y_1x_2y_2, y_1x_2, y_1y_2^2, y_1y_2, y_1, x_2^2, x_2y_2, x_2, y_2^2, y_2, 1$. This is because $p(\mathbf{y})$ is a product of two quadratic forms, neither of which contains terms that are quadratic in a single element. This condition tightly constrains the form of polynomials that arise from multiple motions.

To enforce this constraint imposed by the Q matrices, we define the following special Veronese map. This map does not generate monomials associated to the zero terms in (7)⁴.

Definition 4 (Perspective Veronese Map) Given K independent motions, the *perspective Veronese map* of degree $(2K)$, μ_{2K} , is defined as

$$\mu_{2K} : \mathbf{y} = [x_1, y_1, x_2, y_2, z]^T \mapsto [x_1^{n_1} y_1^{n_2} x_2^{n_3} y_2^{n_4} z^{n_5}]^T, \quad (18)$$

where $n_1 + n_2 \leq K$, $n_3 + n_4 \leq K$, and $n_1 + \dots + n_5 = 2K$.

Finally, we are now ready to estimate a maximal set of $(2K)$ -th degree linearly independent polynomials from the data \mathcal{Y} of K motions. The process is conceptually identical to the one first demonstrated by Vidal et al. (2005) but with the Veronese map ν replaced by the new perspective Veronese map μ .

Define a data matrix V_{2K} that collects all the embedded samples of $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$:

$$V_{2K}(\mathcal{Y}) = (\mu_{2K}(\mathbf{y}_1), \mu_{2K}(\mathbf{y}_2), \dots, \mu_{2K}(\mathbf{y}_N)). \quad (19)$$

Suppose that enough samples are drawn to cover the union of QFM’s and QHM’s. Then, any vector \mathbf{c} in the left null space of V_{2K} corresponds to the coefficients of a $(2K)$ -th degree polynomial $p(\mathbf{y}) = \mathbf{c}^T \mu_{2K}(\mathbf{y})$ that vanishes on \mathcal{Y} . If $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ is a basis for the null space, then the polynomial subspace $I_{2K}(\mathcal{Y})$ has the following basis:

$$I_{2K}(\mathcal{Y}) = \langle \mathbf{c}_1^T \mu_{2K}(\mathbf{y}), \mathbf{c}_2^T \mu_{2K}(\mathbf{y}), \dots, \mathbf{c}_k^T \mu_{2K}(\mathbf{y}) \rangle. \quad (20)$$

It is easy to see that the polynomials in (20) are also linearly independent.

3 Robust Estimation of Segmentation Polynomials

In the previous section, we demonstrated that an HPC $I_{2K}(\mathcal{Y})$ can be recovered from the null space of the data matrix $V_{2K}(\mathcal{Y})$. Nevertheless, when \mathcal{Y} is contaminated with data noise and outliers, $V_{2K}(\mathcal{Y})$ will, in general, be full rank, and will thus have a trivial null space. In this section, we investigate robust estimation of vanishing polynomials in $I_{2K}(\mathcal{Y})$ in the presence of both data noise and outliers.

3.1 Segmentation Polynomial

When \mathcal{Y} is perturbed by moderate data noise, the estimation of its null space becomes a statistical problem. A common

method for estimating the original null space in the presence of noise is *singular value decomposition* (SVD). That is, if the dimension k of the null space is known, then the null space is approximated by a subspace spanned by the eigenvectors corresponding to the k smallest singular values.

The null space dimension k corresponds to the number of linearly independent vanishing polynomials in $I_{2K}(\mathcal{Y})$, which in turn depends on the mixture of rigid body motions and planar homographies. Recently, a closed-form solution was discovered for the number of linearly independent vanishing polynomials for subspace arrangements (Ma et al. 2008). However, to the best of our knowledge, there is still no theoretical solution for k when the algebraic set is a union of quadratic manifolds. Empirical estimation of k from the singular values of $V_{2K}(\mathcal{Y})$ is also known to be not stable in the presence of noise.

To address this issue, we propose a novel approach to segment mixture motions using a single, well-chosen vanishing polynomial, which we shall call a *segmentation polynomial*. To this end, we show that we do not need all of the linearly independent vanishing polynomials to reliably segment image correspondences from different motion models. In this section, we study how to robustly estimate the segmentation polynomial from noisy data samples. The segmentation algorithm based on the segmentation polynomial is studied in Sect. 4.

The coefficients of a polynomial that most likely vanishes on \mathcal{Y} , in a least-squared sense, can be found as the singular vector corresponding to the minimal singular value of V_{2K} . This is equivalent to finding the polynomial coefficient vector \mathbf{c} that minimizes:

$$\mathbf{c}^* = \underset{\mathbf{c}}{\operatorname{argmin}} \mathbf{c}^T \Sigma(\mathcal{Y}) \mathbf{c}, \quad (21)$$

where $\Sigma(\mathcal{Y}) = \sum_{\mathbf{y} \in \mathcal{Y}} \mu_{2K}(\mathbf{y}) \mu_{2K}(\mathbf{y})^T = V_{2K} V_{2K}^T$.

However, in the presence of noise, this polynomial is *not* the optimal polynomial for the purposes of *segmentation*. Inspired by Taubin (1991), we consider minimizing an objective function known as the *Sampson distance*.

Definition 5 (Jacobian of polynomials) Let $Q = [q_1, \dots, q_M]^T$ be a set of (linearly independent) polynomials of five variables $\{x_1, y_1, x_2, y_2, z\}$. Then the *Jacobian* of Q is:

$$\mathcal{J}(Q) = \begin{bmatrix} \frac{\partial q_1}{\partial x_1} & \frac{\partial q_1}{\partial y_1} & \dots & \frac{\partial q_1}{\partial z} \\ \frac{\partial q_2}{\partial x_1} & \frac{\partial q_2}{\partial y_1} & \dots & \frac{\partial q_2}{\partial z} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial q_M}{\partial x_1} & \frac{\partial q_M}{\partial y_1} & \dots & \frac{\partial q_M}{\partial z} \end{bmatrix}. \quad (22)$$

Particularly, for a vanishing polynomial $p(\mathbf{y}) = \mu_{2K}(\mathbf{y})^T \mathbf{c}$, its Jacobian is:

$$\mathcal{J}(p(\mathbf{y})) = \mathcal{J}(\mu_{2K}(\mathbf{y}))^T \mathbf{c}. \quad (23)$$

⁴The constraint on the possible monomials in this map is a necessary but *insufficient* condition for enforcing all of the zeroes in the Q matrices: It is possible to find matrices with nonzero entries on the off diagonal entries that still satisfy the constraint.

Assume a point \hat{y} from an algebraic set \mathcal{A} defined by a set of polynomials Q , i.e., $Q(\hat{y}) = 0$. Denote y to be a noisy version of \hat{y} , which is close to \mathcal{A} :

$$y = \hat{y} + n. \tag{24}$$

Using Taylor expansion, the value of Q evaluated at \hat{y} is given by

$$Q(\hat{y}) = Q(y) + \mathcal{J}(Q(y))(\hat{y} - y) + O(\|\hat{y} - y\|^2). \tag{25}$$

Ignoring the higher order terms, and applying $Q(\hat{y}) = 0$, we obtain the approximate residual.

$$y - \hat{y} \approx (\mathcal{J}(Q(y))\mathcal{J}(Q(y)))^{-1} \mathcal{J}(Q(y))^T Q(y) \tag{26}$$

Definition 6 (Sampson distance) The *Sampson distance* between y and \hat{y} is defined by:

$$\|\hat{y} - y\| \approx \sqrt{Q(y)^T (\mathcal{J}(Q(y))\mathcal{J}(Q(y))^T)^\dagger Q(y)} \in \mathbb{R}, \tag{27}$$

where A^\dagger for a matrix A denotes the Moore-Penrose pseudo-inverse.

Thus, given a set of noisy samples \mathcal{Y} , minimizing the average squared Sampson distance

$$\frac{1}{|\mathcal{Y}|} \sum_{y \in \mathcal{Y}} Q(y)^T (\mathcal{J}(Q(y))\mathcal{J}(Q(y))^T)^\dagger Q(y) \tag{28}$$

leads to optimal estimation of the vanishing polynomials Q (Taubin 1991).

Solving the above minimization problem (28) typically requires nonlinear optimization via iterative gradient-descent techniques. However, under mild assumptions, the optimal polynomials that minimize the Sampson distance (28) can be solved linearly, using a Rayleigh-quotient type criterion (Ma et al. 2008; Yang et al. 2005).

Let $p(y) = \prod_{j=1}^n (y^T Q_j y)$ be a vanishing polynomial for a mixture of n motion models, and let y_1 be a joint image sample from the first motion model (i.e. $y_1^T Q_1 y_1 = 0$). From (5) we obtain:

$$\mathcal{J}(p(y_1)) = \left(\prod_{j=2}^n y_1^T Q_j y_1 \right) (2Q_1 y_1).$$

If y_1 does not lie in any intersection of multiple motion models, then $(y_1^T Q_j y_1 \neq 0), 2 \leq j \leq n$. Thus $\|\mathcal{J}(p(y_1))\|$ will be proportional to the (algebraic) distance between y_1 and all other quadratic manifolds. As a result, a good “segmentation polynomial” should simultaneously minimize the distance between a joint image sample y and its quadratic manifold, while keeping y far from all other quadratic manifolds.

Definition 7 (Segmentation Polynomial) A *segmentation polynomial* in $I_{2K}(\mathcal{Y})$ is a vanishing polynomial with coefficient vector c^* obtained from the following minimization:⁵

$$c^* = \operatorname{argmin}_c \frac{c^T \Sigma(\mathcal{Y})c}{c^T W(\mathcal{Y})c}, \tag{29}$$

where

$$\Sigma(\mathcal{Y}) \doteq \sum_{y \in \mathcal{Y}} \mu_{2K}(y)\mu_{2K}(y)^T, \tag{30}$$

and

$$W(\mathcal{Y}) \doteq \sum_{y \in \mathcal{Y}} \mathcal{J}(\mu_{2K}(y))\mathcal{J}(\mu_{2K}(y))^T. \tag{31}$$

In Definition 7, the matrix Σ measures the square of the fitting polynomial evaluated at each sample, which can be regarded as a squared “algebraic distance” between samples and the quadratic manifolds they lie on. Similarly, matrix W measures the squared norm of the derivative of the polynomial evaluated at each sample, which can be regarded as a squared “distance” between samples and all of the other quadratic manifolds in the multi-motion model. This ratio is similar to the Rayleigh quotient used in *linear discriminant analysis* (LDA). The minimization of the Rayleigh quotient only requires that Σ and W be real, symmetric, positive semidefinite matrices. Thus, the vector c^* that minimizes this ratio is the minimal generalized eigenvector of Σ and W .

3.2 Robust Estimation via Influence Functions

In this subsection, we further consider robust estimation of segmentation polynomials when the data are contaminated by a moderate amount of outliers. It is known that the breakdown point for estimating vanishing polynomials using either SVD (21) or the Rayleigh quotient (29) is 0%. That is, a single outlier with a large magnitude may arbitrarily perturb the singular values and the corresponding subspaces (Jolliffe 2002). As a result, the estimated segmentation polynomial $p(y) = \mu_{2K}(y)^T c^*$ will not vanish on the valid inlying samples in \mathcal{Y} .

In this paper, we use a set of vanishing polynomials as a global constraint to represent mixture motions. Since the coefficients of the polynomials are estimated via recovering eigenvectors of the data matrix $V_{2K}(\mathcal{Y})$, the outliers that affect the polynomials can be rejected by robust estimation of the eigenvalues and eigenvectors. In robust statistics, such methods are referred to as *robust principal component analysis* (Robust PCA), and they are rooted on either

⁵To prevent the denominator of the quotient in (29) from being zero, we add a regularization term ϵI to $W(\mathcal{Y})$, where ϵ is a small positive number and I is the identity matrix.

probability or influence based approaches (Jolliffe 2002). In this paper, we choose to present a means of robustly estimating segmentation polynomials using influence functions. For detailed comparisons between influence functions and the other two approaches, the reader is referred to Yang et al. (2006) and Yang (2006).

As mentioned in Sect. 3.1, the vector of coefficients of a segmentation polynomial $p(\mathbf{y})$ is the generalized eigenvector \mathbf{c} that minimizes the Rayleigh quotient (29). For a particular sample \mathbf{y}_k , a new coefficient vector $\mathbf{c}^{(k)}$ can be estimated using all but the k -th sample:

$$\mathbf{c}^{(k)} \doteq \underset{\mathbf{c}}{\operatorname{argmin}} \frac{\mathbf{c}^T \Sigma^{(k)} \mathbf{c}}{\mathbf{c}^T W^{(k)} \mathbf{c}}, \tag{32}$$

where $\Sigma^{(k)} \doteq \Sigma(\mathcal{Y} \setminus \{\mathbf{y}_k\})$, and $W^{(k)} \doteq W(\mathcal{Y} \setminus \{\mathbf{y}_k\})$. The influence of \mathbf{y}_k on \mathbf{c} and hence $p(\mathbf{y})$ is then defined as the difference between the vector \mathbf{c} and the vector $\mathbf{c}^{(k)}$:

$$I(\mathbf{y}_k) \doteq \langle \mathbf{c}, \mathbf{c}^{(k)} \rangle, \tag{33}$$

where $\langle \cdot, \cdot \rangle$ denotes the subspace angle difference between two subspaces/vectors. In robust statistics, $I(\mathbf{y}_k)$ is called the *sample influence function* (Hampel et al. 1986). Given an influence function, the samples with the highest influence values will be rejected as “outliers”, and will not be used for the estimation of the segmentation polynomial.⁶ The empirical breakdown point of the sample influence function for robust PCA is typically above 30–50%.

3.3 Estimation of Outlier Percentage

Given the estimation of sample influence values for all image correspondences, the outlier rejection problem is still not completely solved. In general, we do not know how many image features are outliers that need to be rejected. In this subsection, we propose a means to estimate the *outlier percentage*. Based on this estimate, the image features with the highest influence will be rejected as outliers. Note that the rejection does not require a segmentation of the data into correct individual motions. The issue of segmentation will be addressed in the next section.

We estimate the outlier percentage as the smallest portion of the total sample set such that the maximal Sampson distance (27) from the remaining samples is smaller than a given residual threshold τ . Equivalently, we reject samples that have large Sampson distance to the union of QFM’s and

QHM’s, and identify those that have small Sampson distance as inliers. The distance threshold τ can be seen as the variation of the noise of the inlying noisy data. The use of τ also helps us to conduct a fair comparison with other robust statistical methods such as RANSAC, where the residual threshold is popularly used to reject outliers. The complete algorithm for robust estimation of the segmentation polynomial is shown as Algorithm 1.

Algorithm 1 (Robust Segmentation Polynomial)

Given a set of joint image samples $\mathcal{Y} = \{\mathbf{y}_k = [x_1, y_1, x_2, y_2, z]^T \in \mathbb{R}^5\}_{k=1}^N$ that are associated with n motions from two perspective images and a Sampson distance threshold τ ,

- 1: Compute polynomial coefficients $\mathbf{c}(\mathcal{Y}) = \min_{\hat{\mathbf{c}}} \frac{\hat{\mathbf{c}}^T \Sigma(\mathcal{Y}) \hat{\mathbf{c}}}{\hat{\mathbf{c}}^T W(\mathcal{Y}) \hat{\mathbf{c}}}$.
 - 2: **for all** $\mathbf{y}_k \in \mathcal{Y}$ **do**
 - 3: Compute $\mathbf{c}^{(k)} = \min_{\hat{\mathbf{c}}} \frac{\hat{\mathbf{c}}^T \Sigma^{(k)} \hat{\mathbf{c}}}{\hat{\mathbf{c}}^T W^{(k)} \hat{\mathbf{c}}}$.
 - 4: Compute influence function $I(\mathbf{y}_k) = \langle \mathbf{c}, \mathbf{c}^{(k)} \rangle$.
 - 5: **end for**
 - 6: **for outlier percentage** $r = 0$ **to** 50 **do**
 - 7: Let outlier set \mathcal{Y}^O be the $r\%$ of samples with largest influence $I(\mathbf{y})$, and inlier set $\mathcal{Y}^I = \mathcal{Y} \setminus \mathcal{Y}^O$.
 - 8: Compute $\Sigma(\mathcal{Y}^I)$, $W(\mathcal{Y}^I)$, and $\mathbf{c}(\mathcal{Y}^I)$.
 - 9: Compute $p(\mathbf{y}) = \mathbf{c}(\mathcal{Y}^I)^T \mu_{2n}(\mathbf{y})$, Jacobian $\mathcal{J}(p(\mathbf{y}))$ and Hessian $H_p(\mathbf{y})$ for all $\mathbf{y} \in \mathcal{Y}^I$.
 - 10: Compute maximal Sampson distance:

$$d^* = \max_{\mathbf{y} \in \mathcal{Y}^I} \sqrt{p(\mathbf{y})^T (\mathcal{J}(p(\mathbf{y})) \mathcal{J}(p(\mathbf{y}))^T)^\dagger p(\mathbf{y})}.$$
 - 11: **if** $d^* \leq \tau$ **then**
 - 12: **break**
 - 13: **end if**
 - 14: **end for**
 - 15: **Output:** $\mathcal{Y} = \mathcal{Y}^I \cup \mathcal{Y}^O$, $p(\mathbf{y})$, $\mathcal{J}(p)$, and H_p .
-

4 Robust Algebraic Segmentation of Quadratic Manifolds

We now show how to use a segmentation polynomial $p(\mathbf{y}) = \mathbf{c}^T \mu_{2K}(\mathbf{y})$ to segment joint image samples \mathcal{Y} into clusters corresponding to different motion models. Because our approach does not rely on the specific form of the Q matrices, the proposed algorithm is able to segment data drawn from a mixture of general quadratic manifolds. Thus, for the following analysis, we assume that the data lie in a D -dimensional space, and their respective quadratic manifolds have dimensions d_1, \dots, d_n (for joint imagery data in our case, $D = 5$ and each d_i will be either 2 or 4).

⁶The major drawback of the sample influence function is the speed. Suppose we have N samples. We then need to compute LDA $N + 1$ times in order to evaluate the influence values of the N samples. In light of this problem, several formulas have been proposed to approximate the function in closed form, called *theoretical influence functions*. For clarity, we only use sample influence functions in this paper. The interested reader is referred to Ma et al. (2008) for further discussion.

As stated before, the estimation of the number of linearly independent vanishing polynomials on $I_{2K}(\mathcal{Y})$ is ill-conditioned. However, without loss of generality, we assume that there are s linearly independent vanishing polynomials on $I_{2K}(\mathcal{Y})$. We notice that an arbitrary polynomial coefficient vector \mathbf{c} is a linear combination of s \mathbf{c}_σ 's, and hence may not correspond to a factorable polynomial. Therefore, such vanishing polynomials are linear combinations of factorable ones:

$$p(\mathbf{y}) = \sum_{k=1}^s a_k p_k(\mathbf{y}) = \sum_{k=1}^s a_k \prod_{i=1}^n \mathbf{y}^T Q_{i\sigma_i(k)} \mathbf{y}. \tag{34}$$

To avoid the difficulty of polynomial factorization, we instead consider the first and second derivatives of the fitting polynomials $\{p(\mathbf{y})\}$. Below we list some of the relevant properties of the derivatives.

Theorem 4 (Derivative of the Fitting Polynomials) *Let $p(\mathbf{y}) = \sum_{k=1}^s a_k \prod_{i=1}^n \mathbf{y}^T Q_{i\sigma_i(k)} \mathbf{y}$ be any multivariate polynomial that vanishes on \mathcal{Y} . If $\mathbf{y}_1 \in \mathcal{Y}$ belongs to the quadratic manifold S_l defined by $\{Q_{lu}\}_{u=1}^{r_l}$, then $\mathcal{J}(p(\mathbf{y}_1)) \in \mathbb{R}^D$, the derivative of $p(\mathbf{y})$ at \mathbf{y}_1 , is given by:*

$$\mathcal{J}(p(\mathbf{y}_1)) = \sum_{u=1}^{r_l} 2\alpha_u(\mathbf{y}_1) Q_{lu} \mathbf{y}_1, \tag{35}$$

where $\alpha_u(\mathbf{y}_1)$ are scalar functions of \mathbf{y} that contain polynomial factors from other manifolds. If \mathbf{y}_1 is on the intersection of more than one manifold, then $\mathcal{J}(p(\mathbf{y}_1)) = 0$.

Proof Without loss of generality, we assume \mathbf{y}_1 is in the first quadratic manifold S_1 , that is $l = 1$. Taking the gradient of $p(\mathbf{y})$ in (34),

$$\begin{aligned} \mathcal{J}(p(\mathbf{y})) &= \sum_{k=1}^s a_k \mathcal{J}\left(\prod_{i=1}^n \mathbf{y}^T Q_{i\sigma_i(k)} \mathbf{y}\right) \\ &= \sum_{k=1}^s a_k \sum_{i=1}^n \left(2Q_{i\sigma_i(k)} \mathbf{y} \prod_{j \neq i} \mathbf{y}^T Q_{j\sigma_j(k)} \mathbf{y}\right). \end{aligned} \tag{36}$$

Since \mathbf{y}_1 is in S_1 , the scalar $\prod_{j \neq i} \mathbf{y}_1^T Q_{j\sigma_j(k)} \mathbf{y}_1 = 0$ for all $i = 2, \dots, n$. Therefore,

$$\begin{aligned} \mathcal{J}(p(\mathbf{y}))|_{\mathbf{y}_1} &= \sum_{k=1}^s \left(2a_k \prod_{j \neq 1} \mathbf{y}_1^T Q_{j\sigma_j(k)} \mathbf{y}_1\right) Q_{1\sigma_1(k)} \mathbf{y}_1 \\ &= \sum_{u=1}^{r_1} 2\alpha_u(\mathbf{y}_1) Q_{1u} \mathbf{y}_1, \end{aligned}$$

where $\alpha_u(\mathbf{y}_1) = \sum_{\sigma_1(k)=u} a_k \prod_{j \neq 1} \mathbf{y}_1^T Q_{j\sigma_j(k)} \mathbf{y}_1$. The terms $\alpha_u(\mathbf{y}_1)$ are scalar functions that contain polynomial factors

from other manifolds. Obviously, when \mathbf{y}_1 is at the intersection of more than one manifold, then in (36), all of the terms $\prod_{j \neq i} \mathbf{y}_1^T Q_{j\sigma_j(k)} \mathbf{y}_1 = 0$, and hence $\mathcal{J}(p(\mathbf{y}_1)) = 0$. \square

Theorem 5 (Surface Normals from the Derivatives) *Let $\mathbf{y} \in \mathcal{Y}$ be a general point that lies exclusively in a single d -dimensional manifold S . Then the matrix*

$$\mathcal{J}(P(\mathbf{y})) \doteq [\mathcal{J}(p_1(\mathbf{y})), \dots, \mathcal{J}(p_s(\mathbf{y}))] \in \mathbb{R}^{D \times s}$$

has rank $r = D - d$. Let the singular value decomposition (SVD) of $\mathcal{J}(P(\mathbf{y}))$ be $\mathcal{J}(P(\mathbf{y})) = U \Sigma V^T$ with U and V being orthogonal matrices and Σ a diagonal matrix. The first r columns of U give a set of orthonormal vectors $\mathbf{n}_1(\mathbf{y}), \dots, \mathbf{n}_r(\mathbf{y})$ to S at \mathbf{y} .

Proof Please refer to Lang (2002). \square

Theorems 4 and 5 allow one to compute the normal vectors to the manifold from the fitting polynomials. If a mixture manifold model consists of only linear subspaces, their normal vectors are indeed sufficient to segment the mixture data, since the normal vectors are invariant for each subspace. This constraint is explicitly harnessed in *generalized principal component analysis* (GPCA) for segmenting mixture subspace models (Vidal et al. 2005).

In the following example, we show that under a special condition, the normal vectors are also sufficient to segment quadratic manifolds.

Example 3 (Segmentation of Quadratic Surfaces with Different Dimensions) Assume all of the quadratic manifolds in the mixture manifold have different dimensions. Then the manifold normals estimated by $\mathcal{J}(P(\mathbf{y}))$ are not a globally invariant, as the normal vectors vary at different locations of a quadratic surface. However, in theory, $\mathcal{J}(P(\mathbf{y}))$ is still sufficient for segmentation: One can segment the points into manifolds with different dimensions by examining the rank of $\mathcal{J}(P(\mathbf{y}))$ at each point \mathbf{y} , which corresponds to the *codimension* of the manifold. In this case, the method harnesses the invariant manifold dimensions that are different among mixture manifolds.

However, as we have stated, determining the number of fitting polynomials from noisy data is an ill-conditioned problem. Also, for most of the problems we investigate, many of the quadratic manifolds in a mixture have the same dimension. Thus, additional information from higher-order derivatives is needed.

Theorem 6 (Hessian of the Vanishing Polynomials) *Let $p(\mathbf{y}) = \sum_{k=1}^s a_k \prod_{i=1}^n \mathbf{y}^T Q_{i\sigma_i(k)} \mathbf{y}$ be any multivariate polynomial that fits \mathcal{Y} . If $\mathbf{y}_1 \in \mathcal{Y}$ belongs to a quadratic manifold S_l defined by $\{Q_{lj}\}_{u=1}^{r_l}$, then the Hessian of $p(\mathbf{y})$ at \mathbf{y}_1 ,*

$H_p(\mathbf{y}_1)$ is given by:

$$\sum_{u=1}^{r_1} [2\alpha_u(\mathbf{y}_1)Q_{lu} + Q_{lu}\mathbf{y}_1\mathcal{J}(\alpha_u(\mathbf{y}_1))^T + \mathcal{J}(\alpha_u(\mathbf{y}_1))(Q_{lu}\mathbf{y}_1)^T], \tag{37}$$

where $\alpha_u(\mathbf{y}_1)$ are scalar functions of \mathbf{y} that contain polynomial factors from other manifolds.

Proof Without loss of generality, we assume \mathbf{y}_1 is in the first quadratic manifold S_1 ; that is, $l = 1$. We have shown in (36) that

$$\mathcal{J}(p(\mathbf{y})) = \sum_{k=1}^s a_k \sum_{i=1}^n \left(2Q_{i\sigma_i(k)}\mathbf{y} \prod_{j \neq i} \mathbf{y}^T Q_{j\sigma_j(k)}\mathbf{y} \right).$$

Hence, its second order derivative is

$$H_p(\mathbf{y}) = \sum_{k=1}^s a_k \sum_{i=1}^n \left[2 \left(\prod_{j \neq i} \mathbf{y}^T Q_{j\sigma_j(k)}\mathbf{y} \right) Q_{i\sigma_i(k)} + 2Q_{i\sigma_i(k)}\mathbf{y} \mathcal{J} \left(\prod_{j \neq i} \mathbf{y}^T Q_{j\sigma_j(k)}\mathbf{y} \right) \right] \tag{38}$$

Substituting \mathbf{y}_1 in (38), and canceling all terms with $\mathbf{y}_1^T Q_{lu}\mathbf{y}_1 = 0$, we get

$$H_p(\mathbf{y}_1) = \sum_{u=1}^{r_1} [2\alpha_u(\mathbf{y}_1)Q_{lu} + Q_{lu}\mathbf{y}_1\mathcal{J}(\alpha_u(\mathbf{y}_1))^T + \mathcal{J}(\alpha_u(\mathbf{y}_1))(Q_{lu}\mathbf{y}_1)^T]. \tag{39}$$

□

Examining (37), we notice that the first term is indeed the Hessian of the factor for the manifold itself, but the second and third terms depend on the derivatives of factors from other manifolds in the mixture. This prevents us from directly using the Hessians to segment the data to different manifolds.

In order to obtain a common signature for each manifold, we seek to eliminate the second and third terms from (37). These two terms are degenerate matrices formed from the outer product of the normal to the manifold $Q_{lu}\mathbf{y}_1$ and another vector $\mathcal{J}(\alpha_u(\mathbf{y}_1))$. Thus, any vector that is tangent to the manifold at \mathbf{y}_1 annihilates these terms. The tangent vectors to the manifold at \mathbf{y}_1 span the null space of the normal to the manifold.

For segmentation, we examine the contraction of the Hessian by the tangent vectors to the manifold.

Definition 8 (Contraction of Hessians) Let $T(\mathbf{y})$ be the matrix $[t_1, \dots, t_d] \in \mathbb{R}^{D \times d}$, a matrix whose columns are ortho-

normal tangent vectors to a manifold at a point \mathbf{y} . Then for every vanishing polynomial $p(\mathbf{y})$, the contraction of $H_p(\mathbf{y})$ by $T(\mathbf{y})$ is defined to be the symmetric matrix:

$$C(\mathbf{y}) \doteq T(\mathbf{y})^T H_p(\mathbf{y}) T(\mathbf{y}) \in \mathbb{R}^{d \times d}.$$

Theorem 7 (Properties of Contractions) Given a mixture of quadratic manifolds, if a point \mathbf{y} is on a quadratic manifold S defined by $\{Q_u\}_{u=1}^r$, then

$$C(\mathbf{y}) = \sum_{u=1}^r 2\alpha_u(\mathbf{y})T(\mathbf{y})^T Q_u T(\mathbf{y}) \in \mathbb{R}^{d \times d} \tag{40}$$

for some scalars $\alpha_u(\mathbf{y}) \in \mathbb{R}$ that contain polynomial factors from other manifolds.

Proof This is a direct result using Theorem 6 and Definition 8. □

The contraction matrix defined in Definition 8 eliminates the second and third terms in (37). However the contraction matrix is still not common for all points of the same quadratic surface, because the tangent and Hessian matrices of two points on the same surface are different. However, we can consider the intersection of the two tangent spaces at the two points. Without loss of generality, let $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}$ be two points on a mixture of d -dimensional quadratic surfaces. Let $T(\mathbf{y}_1)$ and $T(\mathbf{y}_2)$ be the two tangent spaces at \mathbf{y}_1 and \mathbf{y}_2 , respectively. Their intersection $T(\mathbf{y}_1, \mathbf{y}_2) \doteq T(\mathbf{y}_1) \cap T(\mathbf{y}_2)$ is in general a $(2d - D)$ -dimensional subspace in \mathbb{R}^D , assuming $2d > D$. Every $\mathbf{t} \in T(\mathbf{y}_1, \mathbf{y}_2)$ is a tangent to the surface at both points.

Definition 9 (Mutual Contraction) For every fitting polynomial $p_k(\mathbf{x}), k = 1, \dots, s$, the mutual contraction for $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}$ is the pair of contractions of Hessians at \mathbf{y}_1 and \mathbf{y}_2 with $T(\mathbf{y}_1, \mathbf{y}_2)$:

$$\bar{C}^k(\mathbf{y}_1, \mathbf{y}_2) \doteq T(\mathbf{y}_1, \mathbf{y}_2)^T H_{p_k}(\mathbf{y}_1) T(\mathbf{y}_1, \mathbf{y}_2),$$

$$\bar{C}^k(\mathbf{y}_2, \mathbf{y}_1) \doteq T(\mathbf{y}_1, \mathbf{y}_2)^T H_{p_k}(\mathbf{y}_2) T(\mathbf{y}_1, \mathbf{y}_2).$$

Notice that both $\bar{C}^k(\mathbf{y}_1, \mathbf{y}_2)$ and $\bar{C}^k(\mathbf{y}_2, \mathbf{y}_1)$ are $(2d - D) \times (2d - D)$ symmetric matrices. Since the space of all $n \times n$ symmetric matrices has dimension $n(n + 1)/2$, we define $M \doteq (2d - D)(2d - D + 1)/2$, and treat the matrix \bar{C} as a vector in \mathbb{R}^M . By the assumption that $2d > D$, M is positive. Then we have the following relation between the subspaces spanned by the two sets of mutual contractions matrices:

Theorem 8 (Mutual Contraction Subspace) Suppose $M > D - d > 0$. If $\mathbf{y}_1, \mathbf{y}_2 \in \mathcal{Y}$ both belong to the same quadratic surface, then we have

$$\begin{aligned} &\text{span}\{\bar{C}^1(\mathbf{y}_1, \mathbf{y}_2), \dots, \bar{C}^s(\mathbf{y}_1, \mathbf{y}_2)\} \\ &= \text{span}\{\bar{C}^1(\mathbf{y}_2, \mathbf{y}_1), \dots, \bar{C}^s(\mathbf{y}_2, \mathbf{y}_1)\}, \end{aligned} \tag{41}$$

which is a proper subspace in \mathbb{R}^M . We call it the mutual contraction subspace between \mathbf{y}_1 and \mathbf{y}_2 .

Proof Suppose the quadratic surface is defined by the set of symmetric matrices $A_j, j = 1, \dots, D - d$. Similarly to the proof of Theorem 7, one can show that both sets of matrices span the same subspace as the following r matrices:

$$T(\mathbf{y}_1, \mathbf{y}_2)^T A_j T(\mathbf{y}_1, \mathbf{y}_2), \quad j = 1, \dots, r. \tag{42}$$

By the assumption that $M > D - d$, the subspace is proper. \square

Corollary 1 (Quadratic Hyper-Surfaces) *If all the quadratic manifolds are of $D - 1$ dimension (i.e. $s = 1$), for two points*

\mathbf{y}_1 and \mathbf{y}_2 to be on the same quadratic manifolds S , we must have

$$\bar{C}(\mathbf{y}_1, \mathbf{y}_2) \sim \bar{C}(\mathbf{y}_2, \mathbf{y}_1) \in \mathbb{R}^{(D-2) \times (D-2)}, \tag{43}$$

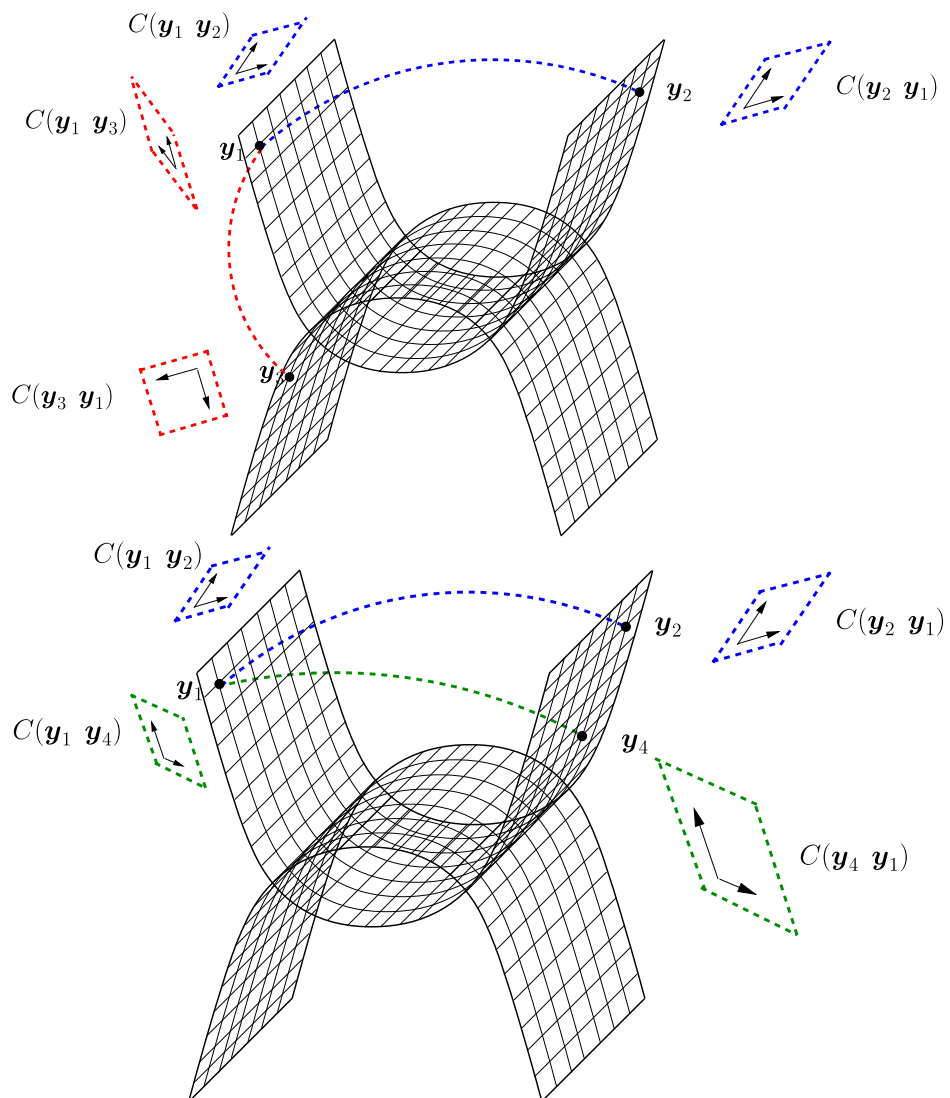
where \sim means “equal up to a nonzero scalar.”

The mutual contraction subspace, unlike the normal vectors for linear subspaces, is *not* globally invariant on the quadratic surfaces. It can be different for different pairs of points ($\mathbf{y}_1, \mathbf{y}_2$) (see Fig. 1). Nevertheless, the mutual contraction subspace gives a very effective necessary condition for segmenting the data points: *If two points belong to the same quadratic manifold, their mutual contraction subspaces are the same.* Thus, if we define

$$W_j = \text{span}\{\bar{C}^1(\mathbf{y}_j, \mathbf{y}_k), \dots, \bar{C}^s(\mathbf{y}_j, \mathbf{y}_k)\} \quad \text{and}$$

$$W_k = \text{span}\{\bar{C}^1(\mathbf{y}_k, \mathbf{y}_j), \dots, \bar{C}^s(\mathbf{y}_k, \mathbf{y}_j)\},$$

Fig. 1 *Top:* A pair of points have the same mutual contraction (up to a scale) if the points are from the same quadratic manifold, and a different mutual contraction otherwise. *Bottom:* Different pairs of points on the same quadratic manifold, in general, have different mutual contractions



we can form a similarity matrix \mathcal{S} with elements

$$\mathcal{S}_{jk} \doteq \begin{cases} \langle W_j, W_k \rangle, & \dim(W_j) = \dim(W_k), \\ 0, & \text{otherwise,} \end{cases} \quad (44)$$

where $\langle \cdot, \cdot \rangle$ is the subspace angle. Then, based on the similarity matrix \mathcal{S} , any spectral clustering algorithm (Ng et al. 2001; Weiss 1999) can be used to segment the data into different quadratic surfaces.

We summarize the results of our derivations in Algorithm 2.

Algorithm 2 (RAS: Robust Algebraic Segmentation)

Given joint image samples $\mathcal{Y} = \{\mathbf{y}_k = [x_1, y_1, x_2, y_2, z]^T \in \mathbb{R}^5\}_{k=1}^N$ from n motions in two perspective images, and a residual threshold τ ,

- 1: Call Algorithm 1 with arguments (\mathcal{Y}, n, τ) to partition $\mathcal{Y} = \mathcal{Y}^I \cup \mathcal{Y}^O$ into an inlier set and an outlier set, and obtain the values of the segmentation polynomial $p(\mathbf{y})$, Jacobian $\mathcal{J}(p(\mathbf{y}))$, and Hessian $H_p(\mathbf{y})$ for all $\mathbf{y} \in \mathcal{Y}^I$.
 - 2: **for all** pair of points $(\mathbf{y}_j, \mathbf{y}_k) \in \mathcal{Y}^I$ **do**
 - 3: Compute $T(\mathbf{y}_j, \mathbf{y}_k) = \mathcal{J}(p(\mathbf{y}_j))^\perp \cap \mathcal{J}(p(\mathbf{y}_k))^\perp$.
 - 4: Compute the mutual contraction matrices

$$\bar{C}_j = T(\mathbf{y}_j, \mathbf{y}_k)^T H_p(\mathbf{y}_j) T(\mathbf{y}_j, \mathbf{y}_k) \quad \text{and}$$

$$\bar{C}_k = T(\mathbf{y}_j, \mathbf{y}_k)^T H_p(\mathbf{y}_k) T(\mathbf{y}_j, \mathbf{y}_k).$$
 - 5: Compute the similarity matrix entry $\mathcal{S}_{jk} = \langle \bar{C}_j, \bar{C}_k \rangle$.
 - 6: **end for**
 - 7: Use a spectral clustering method on \mathcal{S} to cluster \mathcal{Y}^I into n groups.
-

5 Improving RAS via Robust Statistical Methods

Robust algebraic segmentation (RAS) in Algorithm 2 provides an algebraic, noniterative algorithm to segment a mixture of rigid-body and planar motions, and is also capable of rejecting outlying image correspondences using robust segmentation polynomials (Algorithm 1). In this section, we show that RAS complements existing iterative statistical methods, particularly for consensus-based methods such as RANSAC. In particular, we show how to apply RANSAC as a post-processing step to RAS to detect and remove outliers. Supplementing RAS with RANSAC also allows us to utilize both the eight-point algorithm and four-point algorithm in RANSAC to identify individual motions that belong to either general rigid bodies or degenerate planar structures. In this paper, we do not study the problem of model selection between epipolar constraints and homography constraints, so long as the respective images are correctly seg-

mented. The interested reader can refer to Kanatani (2002a) and Schindler and Suter (2005).

The approach proceeds as follows. Given a set of image correspondences known to belong to n motions, we first call RAS with a relaxed outlier rejection parameter τ to partition the data set into n subsets plus an outlier set. We then apply RANSAC to each inlying subset individually to detect and remove outliers.⁷ Within each inlying subset, we define eight joint image samples as the minimum sampling set. In each iteration, RANSAC uses the eight-point algorithm to recover a fundamental matrix F , and calculates its consensus in the group w.r.t. F and a robust threshold τ_F . This process is repeated over a number of trials T . Subsequently, the optimal solution F^* corresponds to the largest consensus among joint image samples within the group, and the samples that do not satisfy the robust threshold τ_F are rejected as outliers.

Next, because a homography is a degenerate epipolar relation, we explicitly use the four-point algorithm to verify if the samples that satisfy the above epipolar constraint are degenerate. We apply the four-point algorithm with a threshold τ_H to all four-point subsets of those eight correspondences corresponding to F^* to obtain a set of candidate homography matrices H 's, and recalculate the consensus for each H . If the maximum consensus is a significant portion of the maximum consensus for F^* (e.g., two-thirds), we label this group as a planar motion, otherwise the group is labeled as a general rigid-body motion. Any feature correspondence that does not adequately satisfy the chosen motion model is labeled an outlier.

The combination of RAS and RANSAC dramatically reduces the complexity of RANSAC applied to mixture models. In contrast, it was shown by Yang et al. (2006) and Schindler and Suter (2005) that the complexity of RANSAC grows exponentially with the number of mixture models, mainly because the percentage of outliers w.r.t. a single model is the sum of the true outliers plus all the inliers from the other mixture models (please refer to Fig. 5 in Sect. 6 for a simulated comparison).

6 Simulation

In this section, we validate the performance of RAS using synthetic image data. We have constructed four types of dynamic scenes for our simulations, referred to as “3F;” “2F+1H;” “1F+2H;” and “3H;” respectively. The numbers before “F” and “H” specify the number of general rigid and planar objects in the given scene, respectively. Thus, these

⁷In our implementation, RANSAC neither reassigns samples of an inlying subset to any other inlying subsets, nor reassign samples in the outlying subset to any inlying subset.

scenes all contain three motions, but each scene has different combinations of general rigid-body and planar motions. In a given trial, three rigid objects of the specified type are generated in space with 150 feature points on each object. In each trial, we randomly vary the location of feature points on each rigid object/plane and location of rigid objects/planes in the scene, the motion of rigid objects/planes and the camera between the two views, the noise perturbation of feature points in the scene, and the location of outlier feature points (if any) in the scene. The three objects all undergo different rigid-body motions and then are projected onto camera frames via perspective projection. Thus the joint image samples from the two image views lie on a mixture of QFMs and QHMs. Two examples of our synthetic dynamic scenes can be seen in Fig. 2.

For scenes with two or more planar structures, we simulate the presence of two walls. Thus, we force two of the planar structures to remain stationary relative to each other between the camera views. This introduces a modeling ambiguity, as the joint image samples from these two structures can be fit either by two planar homographies or a *single* general rigid motion. However, each algorithm is provided with the correct number of motion models. The ground-truth segmentation assigns the joint image samples from these two planar structures to two different planar homographies.

We compare the performance of RAS with four other methods: a greedy agglomerative approach to subspace segmentation called agglomerative lossy compression (ALC) (Rao et al. 2008), a form of RANSAC that finds motion models one at a time (MLESA) (Torr and Davidson 2003), a form of RANSAC that globally finds multiple motion models via Monte-Carlo sampling (MC-RANSAC) (Schindler and Suter 2005), and the projection-based M-estimators (pbM) (Subbarao and Meer 2006). All implementations of the algorithms were obtained from their respective authors.

6.1 Performance on Noisy Data

We measure the misclassification rate of each algorithm as a function of added image noise. After image correspondences are generated as previously described, each image pixel is perturbed by random noise that ranges from 0 to 2 pixels w.r.t. an image size of roughly 1024×1024 . We conduct 200 trials for each combination of rigid-body motion and noise level, and compute the average misclassification rate w.r.t. to the ground-truth labeling. The results of this simulation are displayed in Fig. 3. RAS obtains a reasonable segmentation of the joint image samples in *all* four scenes. RAS performs exceedingly well in the scenes with planar motions, achieving less than 2% misclassification rate in those three cases. The presence of QHMs in the joint image space generally results in many linearly independent vanishing polynomials. Hence, as these results show, the choice of

the segmentation polynomial is the one that best segments the joint image samples. The highest misclassification rate at two-pixel noise level is 5.5% for the 3F model.

We observe that ALC also obtains good segmentations of 3F, although it has been designed to segment linear affine motions. One possible reason could be that, because QFMs are hypersurfaces in the joint image space, a QFM can be reasonably approximated as a Gaussian blob. Thus, the mixture of three general rigid-body motions in “3F” is well modeled by the mixture of Gaussians implicitly used in ALC. The observation is consistent with the drastic decrease in performance as more planar motions are added to the scene.

As for MLESAC, though it works well for mixtures of pure general rigid motion, the greedy nature of its maximum consensus search causes it to fail to segment multiple planar parts of a single motion (the two “walls” in 1F+2H and 3H). Conversely, MC-RANSAC has some trouble segmenting general rigid-body motions, but its accuracy does increase in scenes with more planar motions. Finally, pbM is unable to obtain adequate segmentations for any of the scenes, even in the absence of noise. Thus, we do not compare our methods with pbM in subsequent simulations and experiments.

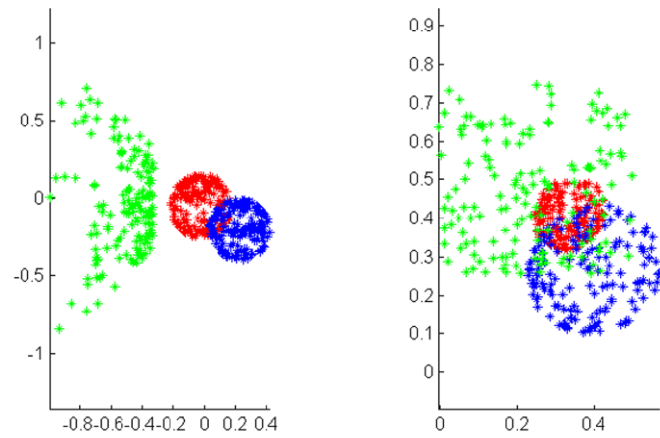
6.2 Performance with Outliers

We further compare the performance of each algorithm on synthetic image data with both pixel noise and outlying image correspondences. First, we test the efficacy of RAS in rejecting outlying image correspondences. We then show the improvement of the segmentation by activating a RANSAC post-processing step of RAS, a method we label as “RAS+RANSAC.”

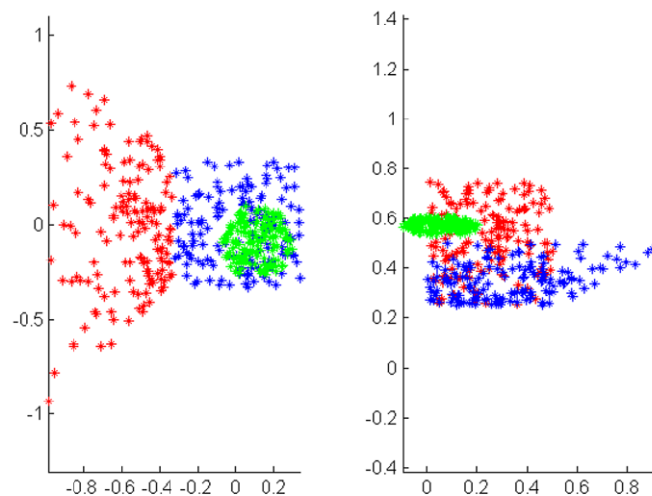
In each trial, 150 feature correspondences are generated from three objects in the same manner as before, and the coordinates of each joint image sample are randomly perturbed by up to one pixel. Then the set of joint image samples is inflated from 0% up to 30% with random correspondences. The algorithms are *not* given the true outlier percentage. We chose the residual threshold τ for RAS and RAS+RANSAC to be 0.22 and 0.25, respectively. The parameters used for RANSAC post-processing are $\tau_F = 0.035$, $\tau_H = 0.04$, 50000 trials, and 2/3 consensus for planar homographies. Figure 4 shows the segmentation accuracy in terms of *false positive rate* (FPR) and *verification rate* (VR).⁸ Once again, we observe that RAS obtains reasonable segmentations for all four scenes, and the accuracy is consistent. With 30% added outliers, RAS achieves 6.33% FPR and 93.8% VR for the 3F model, and 4.9% FPR and 98%

⁸False Positive Rate is the percentage of samples that are either falsely classified as inliers or assigned to the wrong group. Verification Rate is the percentage of samples that are falsely classified as outliers.

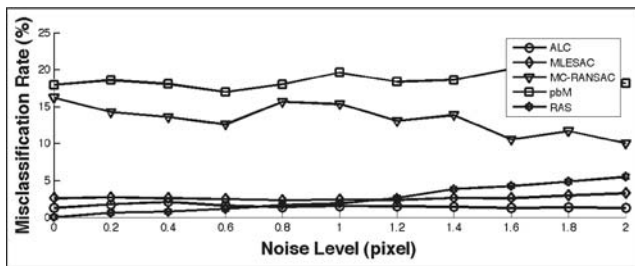
Fig. 2 Two examples of image correspondences of mixed motions (*in color*)



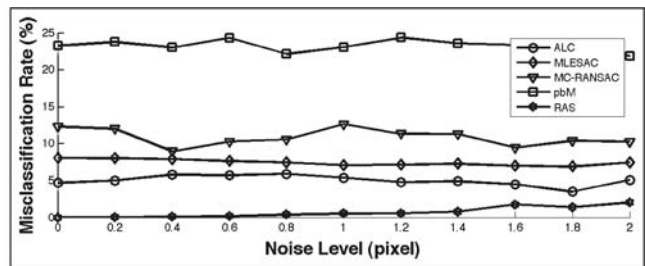
(a) Two perspective images of a (2F + 1H) model. The green planar structure simulates a static wall in space.



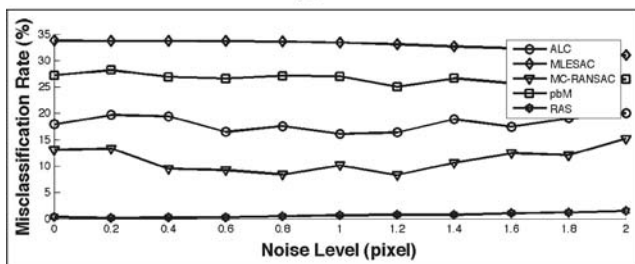
(b) Two perspective images of a (3H) model. The red and blue planar structures simulate two static walls in space.



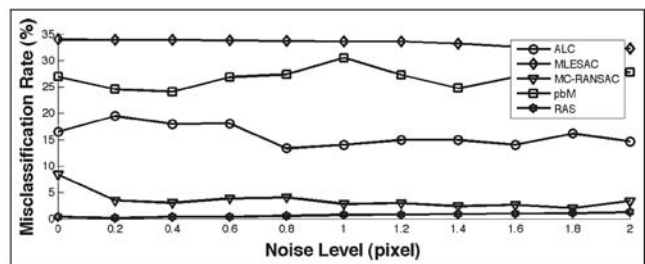
(a) 3F



(b) 2F + 1H



(c) 1F + 2H



(d) 3H

Fig. 3 Comparison of misclassification rates of various algorithms on three rigid bodies in space with different types of motions and added noise

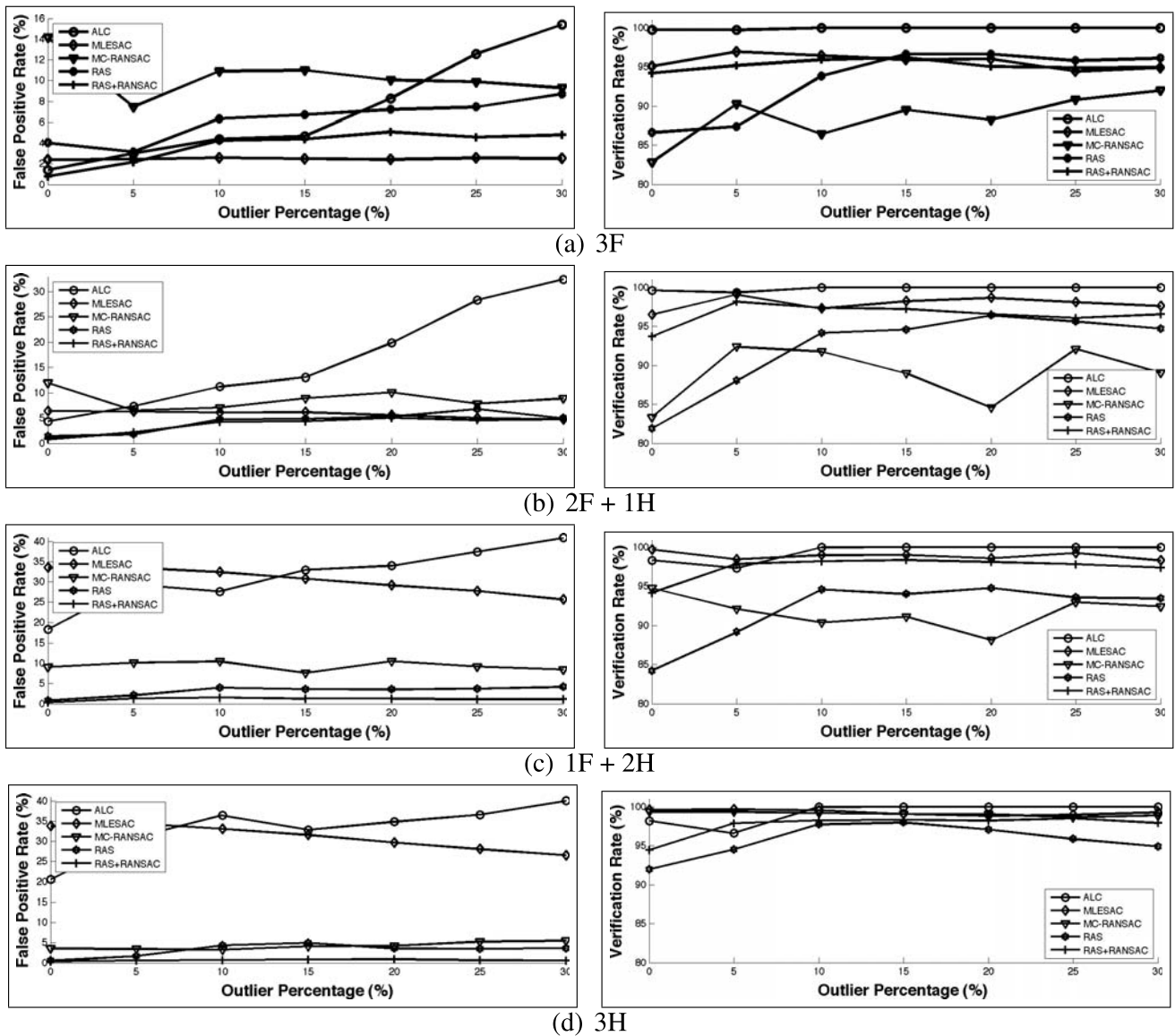


Fig. 4 Performance of various algorithms on four dynamic scenes with different types of motions and synthetic outliers. *Left:* False Positive Rate. *Right:* Verification Rate

VR for the 3H model. Supplementing RAS with RANSAC further improves performance, in terms of both FPR and VR. With 30% added outliers, RAS+RANSAC achieves 4.3% FPR and 95.9% VR for the 3F model, and a meager 0.8% FPR and 98.3% VR for the 3H model.

We compare the performance of RAS and RAS+RANSAC with other methods. ALC can obtain a decent segmentation for the 3F model with no outliers, but does not work well with increased numbers of outliers and/or planar motions. MLESAC obtains good results for the 3F and 2F+1H models, and the performance is independent of the number of outliers. However, just as in the noise case, MLESAC cannot segment multiple planar motions of the same rigid object. Finally, while MC-RANSAC does not fail on

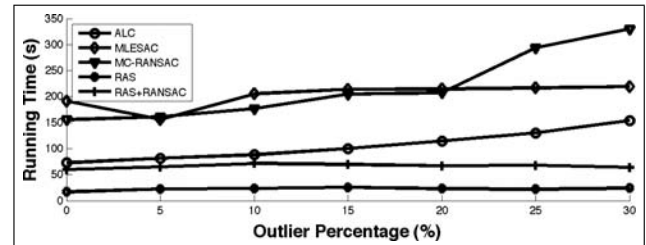


Fig. 5 The average running time of various algorithms on synthetic scenes as a function of added outlier percentage

any of the models, its performance, in both FPR and VR, is consistently worse than RAS+RANSAC.

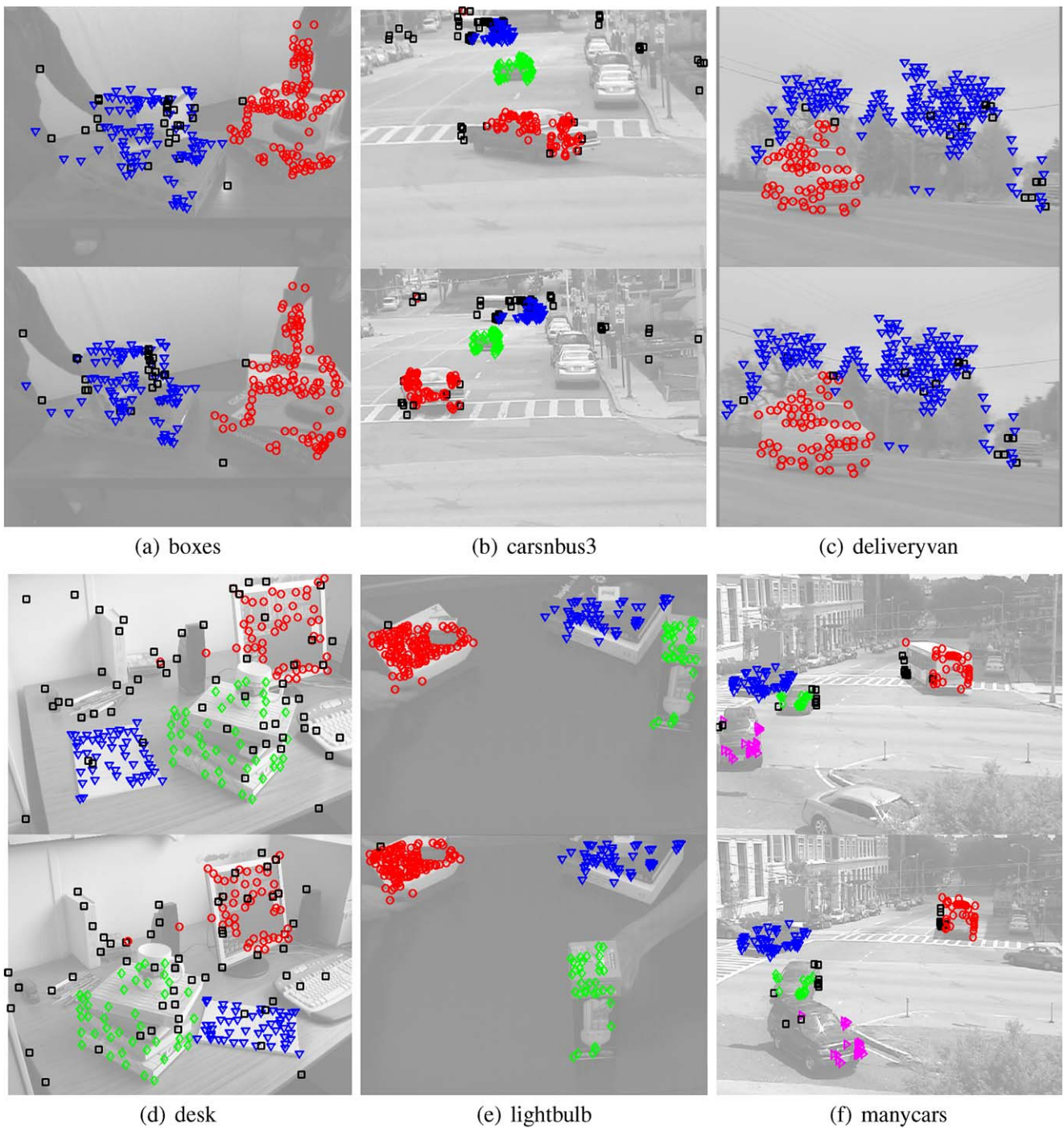


Fig. 6 Segmentation results (*in color*) of RAS+RANSAC on nine real image sequences. Features assigned to different motions are labeled with different colors and shapes. Features classified as outliers are labeled as black squares

We also measure the average running time of each method as a function of outlier percentage. As Fig. 5 shows, both RAS and RAS+RANSAC are considerably more efficient, running at least *two to three times* faster than methods based on RANSAC. In addition, the running times of RAS and RAS+RANSAC scale gracefully with added out-

liers. In summary, both the proposed algorithm, RAS, and its variation, RAS+RANSAC, achieve the best overall segmentation performance in simulations with mixed rigid-body motions, and their running times are only a small fraction of the running times of most random sampling-based algorithms.

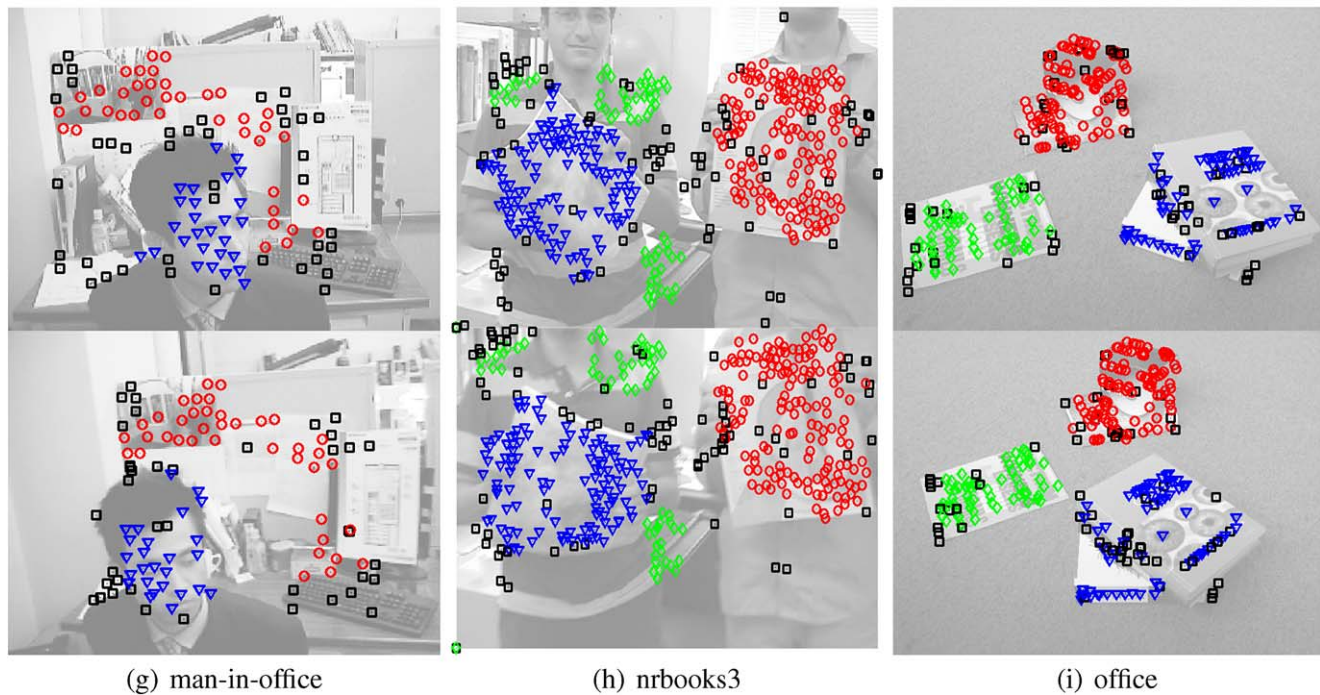


Fig. 6 (Continued)

7 Experiment

In this section we verify the performance of RAS and RAS+RANSAC on the 13 real image sequences listed in Table 1. We obtained the sequences “boxes,” “deliveryvan,” “desk,” “lightbulb,” and “office” from (Schindler and Suter 2005), “man-in-office,” “parking-lot,” and “toys-on-table” from (Sugaya and Kanatani 2003), and “carsnbus3,” “manycars,” and “nrbooks3” from (Tron and Vidal 2007). In addition, we created the sequences “posters-checkerboard” and “posters-keyboard” to mimic the simulation models 3H and 1F+2H, respectively.

For each sequence, we apply our methods RAS and RAS+RANSAC, as well as MLESAC and MC-RANSAC.⁹ For sequences that contain features from more than two image frames, we simply use the first and last image frames to construct joint image samples. All parameters for RAS and RAS+RANSAC are the same as those used in Sect. 6, and parameters for all other methods are tuned to achieve the best segmentation results. The segmentation results for RAS+RANSAC on each sequence are shown in Figs. 6 and 7. The performance results (in terms of FPR and VR) for all methods are listed in Table 2.

RAS outperforms MLESAC on all but one sequence, and MC-RANSAC on eight of the thirteen sequences. RAS+

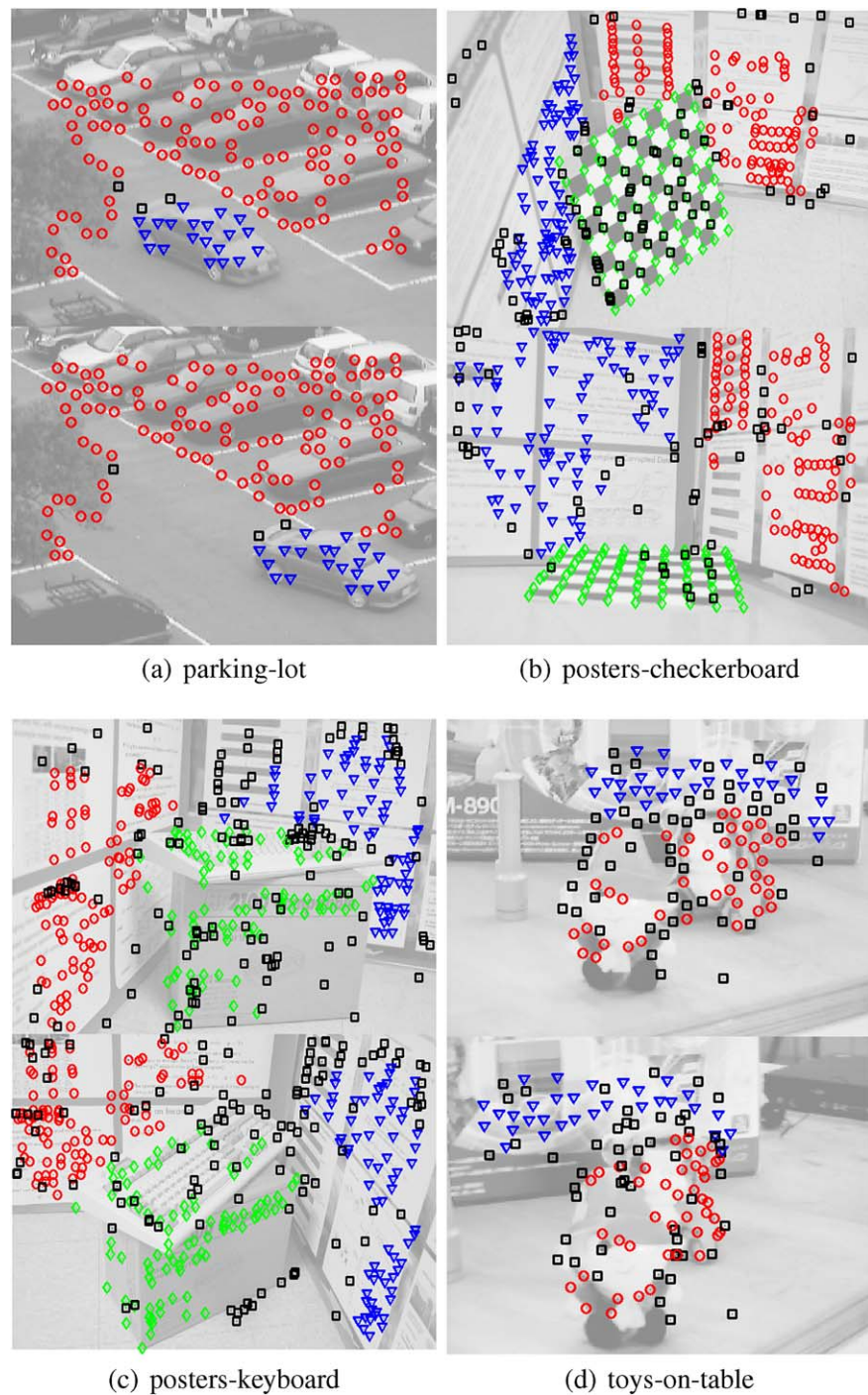
Table 1 Information about image sequences used for experiments

Sequence	Models	# Samples	# Outliers
boxes	2F	115, 121	2
carsnbus3	3F	85, 45, 89	28
deliveryvan	2F	62, 192	0
desk	1F+2H	50, 50, 55	45
lightbulb	3F	51, 121, 33	0
manycars	4F	54, 24, 23, 43	0
man-in-office	2F	16, 57	34
nrbooks3	1F+2H	129, 168, 91	32
office	2F+1H	76, 109, 74	48
parking-lot	2F	19, 117	4
posters-checkerboard	3H	100, 99, 81	99
posters-keyboard	1F+2H	99, 99, 99	99
toys-on-table	1F+1H	49, 42	35

RANSAC further improves the accuracy of RAS, achieving almost perfect segmentation for most of the sequences, even in cases where the segmentation by RAS is poor. The notable exceptions are the sequences from Sugaya and Kanatani (“man-in-office,” “parking-lot,” and “toys-on-table”). These sequences contain multiple image frames that are continuous in time, and were originally designed for affine camera models. In this experiment, only the first and last frames from each of the sequences are used. Hence the segmentation problem becomes much more challenging. In

⁹As ALC is designed to work with sequences with many image frames, we do not compare our methods with it in this section.

Fig. 7 Segmentation results (*in color*) of RAS+RANSAC on four more real image sequences. Features assigned to different motions are labeled with different *colors* and *shapes*. Features classified as outliers are labeled as *black squares*



addition, some of the motions in these sequences have very few tracked correspondences, making it difficult for RAS to find a polynomial that vanishes on all of the joint image data. Nevertheless, RAS+RANSAC still obtains the best segmentation of the methods we tested on these sequences. MC-RANSAC is able to achieve comparable segmentation results on some sequences, but tends to falsely reject many more inliers.

8 Conclusion

We have proposed a general framework for segmentation of mixed rigid-body motions and planar motions, whose features lie on certain quadratic manifolds in the joint image space. We have proved that the joint image of features from K mixture epipolar and homography constraints is uniquely determined by a set of $(2K)$ -th degree vanishing poly-

Table 2 False positive and verification rates for various algorithms on thirteen real image sequences. Best results are in bold

boxes	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	9.24%	0.84%	1.68%	0.84%
VR	36.97%	84.87%	100%	87.39%
carsnbus3	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	45.75%	12.55%	2.83%	1.62%
VR	83.81%	90.28%	97.17%	85.83%
deliveryvan	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	23.23%	10.63%	5.91%	0.39%
VR	97.64%	96.85%	100%	94.09%
desk	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	9.00%	2.50%	3.00%	0.50%
VR	55.50%	93.50%	91.50%	93.50%
lightbulb	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	39.52%	0.00%	0.00%	0.00%
VR	76.19%	82.86%	100%	99.52%
manycars	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	30.56%	22.22%	0.00%	0.00%
VR	90.28%	95.83%	100%	88.89%
man-in-office	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	20.56%	34.58%	20.56%	11.21%
VR	89.72%	95.33%	84.11%	82.24%
nrbooks3	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	12.38%	9.05%	5.48%	0.95%
VR	41.19%	65.48%	94.29%	88.33%
office	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	2.28%	0.33%	10.42%	0.00%
VR	89.59%	90.55%	86.97%	93.49%
parking-lot	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	7.86%	5.00%	3.57%	2.86%
VR	98.57%	96.43%	100%	97.86%
posters-checker	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	20.58%	1.06%	9.23%	0.00%
VR	49.87%	97.36%	70.71%	95.25%
posters-keyboard	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	8.59%	0.25%	10.61%	0.51%
VR	56.06%	83.33%	78.03%	88.13%
toys-on-table	MLESAC	MC-RANSAC	RAS	RAS+RANSAC
FPR	38.10%	38.10%	15.08%	7.94%
VR	91.27%	92.86%	81.75%	77.78%

mials. The membership of the feature points in term of the K motions can be derived from the first and second derivatives of the recovered vanishing polynomials. We note that our current method requires that K is provided as an input. Automatic determination of K , by applying a prior such as spatial continuity, is a topic for future research.

Throughout the paper, we have emphasized the confluence of algebra and statistics in segmentation of mixture geometric models (here quadratic manifolds). The new algebraic algorithm complements traditional robust segmentation methods that are based on statistical inference only. On one hand, we have shown that certain techniques in robust statistics can effectively reject outlying samples in the estimation of segmentation polynomials, and improve the accuracy of segmentation with noisy image samples. On the other hand, the global algebraic algorithm, RAS, can be coupled with the RANSAC, which not only reduces the number of iterations but also improves the final segmentation result.

The efficacy of the proposed algorithm, RAS, has been validated by extensive simulations and experiments on real, public motion sequences. The comparison of RAS with several other robust motion segmentation algorithms (i.e., MLESAC, MC-RANSAC, ALC, and pbM) has shown that RAS and its variation, RAS+RANSAC, clearly outperform the existing algorithms, particularly when a dynamic scene contains multiple motions with different types. The experiment has also demonstrated that the implementation of RAS is two to three times faster than the implementation of the other robust algorithms.

Acknowledgements The authors would like to thank Roberto Tron and Prof. Rene Vidal of Johns Hopkins University, Dr. Konrad Schindler of ETH Zurich, Dr. Raghav Subbarao and Prof. Peter Meer of Rutgers University, Prof. Kenichi Kanatani of Okayama University, and Prof. Philip H. S. Torr of Oxford Brooks University for providing us with code and test sequences. This work is partially supported by NSF CAREER IIS-0347456, ONR YIP N00014-05-1-0633, and ARO MURI W911NF-06-1-0076.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Proof of Theorem 2 One direction of the equalities is trivial. Since I^F is generated by $g^F \in I(S^F)$, hence $I^F \subseteq I(S^F)$. Similarly, $I^H \subseteq I(S^H)$.

To show the other direction, by Hilbert's Nullstellensatz theorem, we only need to prove that I^F and I^H are prime ideals (Lang 2002).

Since $I^F = Rg^F$, i.e., it only has one generator, I^F is a principal ideal. Therefore, I^F is a prime ideal. By the Nullstellensatz theorem, $I^F \supseteq I(S^F)$.

For I^H , we need to show by definition that if $p_1 p_2 \in I^H$, either $p_1 \in I^H$ or $p_2 \in I^H$. Since $p_1 p_2 \in I^H$, $p_1 p_2 = a_1 g_1^H + a_2 g_2^H + a_3 g_3^H$ for some polynomials $a_1, a_2, a_3 \in R$. By the assumption, we also know that g_1, g_2 and g_3 are linearly independent.

Suppose that $p_1 \notin I^H$, hence by polynomial division, $p_1 = b_1 g_1^H + b_2 g_2^H + b_3 g_3^H + b_4$ for some polynomials $b_1, b_2, b_3, b_4 \in R$, and b_4 is linearly independent with g_1, g_2 and g_3 . Therefore,

$$\begin{aligned} p_1 p_2 &= (b_1 g_1^H + b_2 g_2^H + b_3 g_3^H + b_4) p_2 \\ &= a_1 g_1^H + a_2 g_2^H + a_3 g_3^H. \end{aligned} \quad (45)$$

Combining the g_1^H, g_2^H , and g_3^H terms on the two sides of the equality, it is easy to see that p_2 must be of the form $c_1 g_1^H + c_2 g_2^H + c_3 g_3^H$ for some $c_1, c_2, c_3 \in R$. Hence, $p_2 \in I^H$. I^H is also a prime ideal. Hence, $I^H \supseteq I(S^H)$.

In summary, $I^F = I(S^F)$ and $I^H = I(S^H)$. \square

References

- Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (1998). Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of ACM special interest group on management of data*.
- Anandan, P., & Avidan, S. (2000). Integrating local affine into global projective images in the joint image space. In *Proceedings of European conference on computer vision*.
- Ballard, D. (1981). Generalizing the Hough transform to detect arbitrary patterns. *Pattern Recognition*, 13(2), 111–122.
- Bartoli, A. (2001). Piecewise planar segmentation for automatic scene modeling. In *Proceedings of IEEE conference on computer vision and pattern recognition*.
- Campbell, N. (1978). The influence function as an aid in outlier detection in discriminant analysis. *Applied Statistics*, 27(3), 251–258.
- Campbell, N. (1980). Robust procedures in multivariate analysis I: robust covariance estimation. *Applied Statistics*, 29(3), 231–237.
- Costeira, J., & Kanade, T. (1998). A multibody factorization method of independently moving objects. *International Journal on Computer Vision*, 29(3), 159–179.
- Fischler, M., & Bolles, R. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24, 381–85.
- Frahm, J., & Pollefeys, M. (2006). RANSAC for (quasi-) degenerate data (QDEGSAC). In *Proceedings of IEEE conference on computer vision and pattern recognition* (pp. 453–460).
- Gnanadesikan, R., & Kettenring, J. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28(1), 81–124.
- Goh, A., & Vidal, R. (2007). Segmenting motions of different types by unsupervised manifold clustering. In *Proceedings of IEEE conference on computer vision and pattern recognition* (pp. 1–6).
- Goshen, L., Shimshoni, I., Anandan, P., & Keren, D. (2005). Motion recovery by integrating over the joint image manifold. *International Journal on Computer Vision*, 65(3), 131–145.
- Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. (1986). *Robust statistics: the approach based on influence functions*. New York: Wiley.

- Han, M., & Kanade, T. (2000). Reconstruction of a scene with multiple linearly moving objects. In *Proceedings of IEEE conference on computer vision and pattern recognition* (Vol. 2, pp. 542–549).
- Harris, J. (1992). *Algebraic geometry: a first course*. Berlin: Springer.
- Huber, P. (1981). *Robust statistics*. New York: Wiley.
- Jolliffe, I. (2002). *Principal component analysis* (2nd edn.). Berlin: Springer.
- Kanatani, K. (2002a). Estimating the number of independent motions for multibody segmentation. In *Proceedings of Asian conference on computer vision* (pp. 7–12).
- Kanatani, K. (2002b). Motion segmentation by subspace separation: model selection and reliability evaluation. *International Journal of Image and Graphics*, 2(2), 179–197.
- Kanatani, K., & Sugaya, Y. (2003). Multi-stage optimization for multibody motion segmentation. In *Proceedings of the Australia-Japan advanced workshop on computer vision*.
- Lang, S. (2002). *Algebra*. Berlin: Springer.
- Lee, K., Ho, J., & Kriegman, D. (2005). Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5), 684–698.
- Leonardis, A., Bischof, H., & Maver, J. (2002). Multiple eigenspaces. *Pattern Recognition*, 35(11), 2613–2627.
- Ma, Y., Derksen, H., Hong, W., & Wright, J. (2007). Segmentation of multivariate mixed data via lossy coding and compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9), 1546–1562.
- Ma, Y., Yang, A., Derksen, H., & Fossum, R. (2008). Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM Review*, 50(3), 413–458.
- Ng, A., Weiss, Y., & Jordan, M. (2001). On spectral clustering: analysis and an algorithm. In *Proceedings of neural and information processing systems*.
- Ozden, K., Schindler, K., & van Gool, L. (2007). Simultaneous segmentation and reconstruction of monocular image sequences. In *Proceedings of IEEE international conference on computer vision*.
- Rao, S., Yang, A., Wagner, A., & Ma, Y. (2005). Segmentation of hybrid motions via hybrid quadratic surface analysis. In *Proceedings of IEEE international conference on computer vision*.
- Rao, S., Tron, R., Vidal, R., & Ma, Y. (2008). Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories. In *Proceedings of IEEE conference on computer vision and pattern recognition*.
- Rousseeuw, P. (1984). Least median of squares regression. *Journal of American Statistics Association*, 79, 871–880.
- Schindler, K., & Suter, D. (2005). Two-view multibody structure-and-motion with outliers. In *Proceedings of IEEE conference on computer vision and pattern recognition*.
- Scholkopf, B., Smola, A., & Muller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299–1319.
- Shapiro, L. (1995). *Affine analysis of image sequences*. Cambridge: Cambridge University Press.
- Shashua, A., & Levin, A. (2001). Multi-frame infinitesimal motion model for the reconstruction of (dynamic) scenes with multiple linearly moving objects. In *Proceedings of IEEE international conference on computer vision* (Vol. 2, pp. 592–599).
- Souvenir, R., & Pless, R. (2007). Manifold Clustering. In *International conference on computer vision*.
- Stewart, C. (1995). MINPRAN: a new robust estimator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10), 925–938.
- Stewart, C. (1999). Robust parameter estimation in computer vision. *SIAM Review*, 41(3), 513–537.
- Subbarao, R., & Meer, P. (2006). Beyond RANSAC: user independent robust regression. In *CVPR workshop on 25 years of RANSAC*.
- Sugaya, Y., & Kanatani, K. (2003). Outlier removal for motion tracking by subspace separation. *IEICE Transactions on Information and Systems*, E86-D(6), 1095–1102.
- Taubin, G. (1991). Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(11), 1115–1138.
- Tipping, M., & Bishop, C. (1999). Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11, 443–482.
- Tong, W., Tang, C., & Medioni, G. (2004). Simultaneous two-view epipolar geometry estimation and motion segmentation by 4D tensor voting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9), 1167–1184.
- Torr, P. (1998). Geometric motion segmentation and model selection. *Philosophical Transactions of the Royal Society of London*, 356(1740), 1321–1340.
- Torr, P., & Davidson, C. (2003). IMPSAC: synthesis of importance sampling and random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3), 354–364.
- Torr, P., & Zisserman, A. (2000). MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, (78), 138–156.
- Triggs, B. (1995). Matching constraints and the joint image. In *Proceedings of IEEE international conference on computer vision*.
- Tron, R., & Vidal, R. (2007). A benchmark for the comparison of 3-D motion segmentation algorithms. In *Proceedings of IEEE conference on computer vision and pattern recognition* (pp. 1–8).
- Vidal, R., & Hartley, R. (2008). Three-view multibody structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), 214–227.
- Vidal, R., & Ma, Y. (2006). A unified algebraic approach to 2-D and 3-D motion segmentation and estimation. *Journal of Mathematical Imaging and Vision*, 25, 403–421.
- Vidal, R., Ma, Y., & Sastry, S. (2005). Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12), 1–15.
- Vidal, R., Ma, Y., Soatto, S., & Sastry, S. (2006). Two-view multibody structure from motion. *International Journal on Computer Vision*, 68(1), 7–25.
- Wang, H., & Suter, D. (2004). MDPE: a very robust estimator for model fitting and range image segmentation. *International Journal on Computer Vision*, 59(2), 139–166.
- Weiss, Y. (1999). Segmentation using eigenvectors: a unifying view. In *Proceedings of IEEE international conference on computer vision* (pp. 975–982).
- Yan, J., & Pollefeys, M. (2006). A general framework for motion segmentation: independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *Proceedings of European conference on computer vision* (pp. 94–106).
- Yang, A. (2006). *Estimation of subspace arrangements: its algebra and statistics*. PhD Dissertation, University of Illinois at Urbana-Champaign.
- Yang, A., Rao, S., Wagner, A., & Ma, Y. (2005). Segmentation of a piece-wise planar scene from perspective images. In *Proceedings of IEEE conference on computer vision and pattern recognition*.
- Yang, A., Rao, S., & Ma, Y. (2006). Robust statistical estimation and segmentation of multiple subspaces. In *CVPR workshop on 25 years of RANSAC*.