Dynamics Based Robust Motion Segmentation*

Roberto Lublinerman
Dept. of Computer Science and Engineering
The Pennsylvania State University
University Park, PA 16802, USA
rluble@psu.edu

Mario Sznaier Octavia Camps
Dept. of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802, USA
{msznaier,camps}@ee.psu.edu

Abstract

In this paper we consider the problem of segmenting multiple rigid motions using multi–frame point correspondence data. The main idea of the method is to group points according to the complexity of the model required to explain their relative motion. Intuitively, this formalizes the idea that points on the same rigid share more modes of motion (for instance a common translation or rotation) than points on different objects, leading to less complex models. By exploiting results from systems theory, the problem of estimating the complexity of the underlying model is reduced to simply computing the rank of a matrix constructed from the correspondence data. This leads to a simple segmentation algorithm, computationally no more expensive than a sequence of SVDs. Since the proposed method exploits both spatial and temporal constraints, is less sensitive to the effect of noise or outliers than approaches that rely solely on factorizations of the measurements matrix. In addition, the method can also naturally handle "degenerate cases", e.g. cases where the objects partially share motion modes. These results are illustrated using several examples involving both degenerate and non–degenerate cases.

1. Introduction

The problem of 2-D motion segmentation, that is determining the number of moving objects from a sequence of 2-D frames and assigning points to each object, has been object of substantial research in the past decade. Approaches to this problem include temporal differences [10], background substraction [18], and optical flow [1, 2, 3, 4, 7, 17, 25, 26]. Tomasi and Kanade [20] proposed the factorization method to recover 3D motion and geometric structure of a single moving object from 2D images under the assumption of orthographic projection. The method has been extended to paraperspective [14, 15] and perspective [19] projection. Several extensions of these methods have been proposed to deal with non rigid objects [22, 30] and to segment multiple moving objects [5, 6, 8, 11, 21, 24, 9, 27].

To illustrate the basic idea behind these methods, consider N_p points from a rigid structure, tracked over N_F frames. Define the measurements matrix $\mathcal{W}$ as:

$$\mathcal{W} \doteq \begin{bmatrix} \mathbf{p}_{11} & \cdots & \mathbf{p}_{1N_p} \\ \vdots & & \vdots \\ \mathbf{p}_{N_F1} & \cdots & \mathbf{p}_{N_FN_p} \end{bmatrix} \tag{1}$$

where $\mathbf{p}_{i,t} \doteq (u_t^i, v_t^i)^T$ containts the image coordinates of the i^{th} feature at frame t. Under the assumption of an affine camera, it can be shown [5] that $\text{rank}(\mathcal{W}) \leq 4$[1]. Thus, in principle, motion segmentation can be accomplished by factoring the column space of $\mathcal{W}$ into (approximately) four–dimensional subspaces. While this idea leads to computationally very efficient solutions, is fragile to noise and the presence of outliers, that can lead to incorrect subspace assignments. Robustness against these effects has been improved in several recently proposed methods that build upon the idea of factorizing suitable subspaces of $\mathcal{W}$ [29, 23, 28]. However, as we illustrate next, due to the underlying geometrical arguments, approaches based solely on these factorizations have difficulties in disambiguating objects that partially share motion modes. An example of this situation is the propellers of the airplane shown in Figure 1, where all objects share the motion of the centroid, while propellers on the same wing also share the relative rotation.

Denote by $\mathbf{p}_i^{(j)}(k) = [u_i^{(j)}(k)\ v_i^{(j)}(k)]^T$ the position at time k of the image of the point $P_i^{(j)}$ on the j^{th} propeller. Using homogeneous coordinates, it can be easily shown that

*This work was supported in part by AFOSR under grant FA9550–05–1–0437 and NSF, under grants IIS–0117387, ECS–0221562 and ITR–0312558.

[1] $\text{rank}(\mathcal{W}) \leq 3$ in the case of planar motion .

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

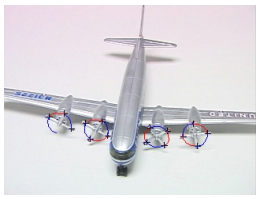

Figure 1. Segmenting the propellers in an airplane.

$\mathcal{W} = \mathcal{M}\mathbf{S}$, where:

$$\mathcal{W} = \begin{bmatrix}\mathbf{W}(1)^T & \dots & \mathbf{W}(N_F)^T\end{bmatrix}^T,$$

$$\mathbf{W}(k) = \begin{bmatrix}\mathbf{p}_1^{(1)}(k) \dots \mathbf{p}_{n_1}^{(1)}(k) \dots \mathbf{p}_1^{(4)}(k) \dots \mathbf{p}_{n_4}^{(4)}(k)\end{bmatrix}$$

$$\mathcal{S} = \begin{bmatrix}\mathbf{S}_1 \\ \mathbf{E}\end{bmatrix}, \ \mathbf{E} = \begin{bmatrix} 1_{1\times n_1} & 0 & 0 & 0 \\ 0 & 1_{1\times n_2} & 0 & 0 \\ 0 & 0 & 1_{1\times n_3} & 0 \\ 0 & 0 & 0 & 1_{1\times n_4}\end{bmatrix}$$

$$\mathbf{S}_1 = \left[\begin{array}{c|c} \mathbf{p}_1^{(1)}(0) \dots \mathbf{p}_{n_2}^{(2)}(0) & 0 \ \dots \ 0 \\ \hline 0 \ \dots \ 0 & \mathbf{p}_1^{(3)}(0) \dots \mathbf{p}_{n_4}^{(4)}(0)\end{array}\right]$$

$$\mathcal{M} = \begin{bmatrix}\mathbf{M}(1) \\ \vdots \\ \mathbf{M}(N_F)\end{bmatrix}, \ \mathbf{M}(k) = \begin{bmatrix} \cos\theta(k) & -\sin\theta(k) \\ \sin\theta(k) & \cos\theta(k) \\ \cos\theta(k) & \sin\theta(k) \\ -\sin\theta(k) & \cos\theta(k) \\ t_x^{(1)}(k) & t_y^{(1)}(k) \\ t_x^{(2)}(k) & t_y^{(2)}(k) \\ t_x^{(3)}(k) & t_y^{(3)}(k) \\ t_x^{(4)}(k) & t_y^{(4)}(k)\end{bmatrix}^T$$

where $t_x^{(j)}, t_y^{(j)}$ denote the position of the centroid of the j^{th} propeller and $\theta(k)$ is the angular displacement of the left wing propellers[2]. It follows that, if at least 2 points in each propeller are tracked and $N_F \geq 6$, then rank $\mathcal{W} = 6$, since $\mathcal{M}$ contains only 6 linearly independent columns. Thus, *any* motion segmentation approach based solely on finding linearly independent subspaces of the column space of $\mathcal{W}$ will fail, since it cannot distinguish this case from the case of two independently moving propellers.

Intuitively, the main difficulty in the example above is that approaches based on properties of $\mathcal{W}$ that are invariant under column permutations, *take into account* only *geometrical constraints, but not temporal ones*. This is due to the fact that any matrix $\hat{\mathcal{W}}$ consisting of a permutation of the columns of $\mathcal{W}$ satisfies the same geometric constraints, but corresponds to different time trajectories. The core idea of this paper is to resolve the ambiguities noted above by exploiting *both* sets of constraints. This is accomplished by grouping points according to the complexity of the model required to explain their spatio–temporal evolution, which, under mild assumptions, can be estimated by computing the ranks of a sequence of matrices obtained directly from the matrix $\mathcal{W}$. In addition, the use of spatio-temporal data allows for improving robustness against measurement noise and outliers. These results are illustrated with several examples involving both degenerate and non-degenerate motion.

[2]Propellers in different wings move with the same angular velocity but in opposite directions.

2. Dynamics based motion segmentation.

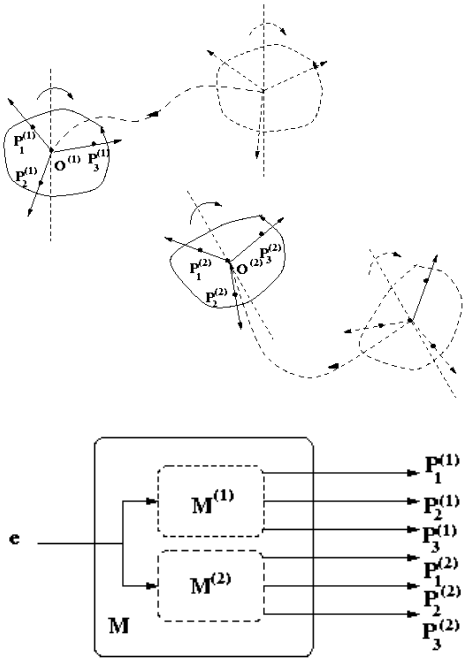

Figure 2. Using models of relative motion to segment. Top: two independently moving objects. Bottom: corresponding dynamical model

To illustrate the main idea of this paper, consider N moving objects and associate to the j^{th} object its centroid $\mathbf{O}^{(j)}$ and an affine basis $\mathbf{b}^{(j)}$, centered at $\mathbf{O}^{(j)}$, defined by three no coplanar vectors $\mathbf{V}_i^{(j)}$ (see Figure 2). Finally, denote by $\mathbf{o}^{(j)}(k)$, $\mathbf{v}_i^{(j)}(k)$ the coordinates of the image of $\mathbf{O}^{(j)}(k)$ and the projections of $\mathbf{V}_i^{(j)}(k)$ onto the image plane at time k, respectively. Given a point $\mathbf{P}_i^{(j)}$ belonging to the j^{th} object, the coordinates at time k of its image $\mathbf{p}_i^{(j)}(k)$ are given by:

$$\mathbf{p}_i^{(j)}(k) = \mathbf{o}^{(j)}(k) + \alpha_i^{(j)}\mathbf{v}_1^{(j)}(k) + \beta_i^{(j)}\mathbf{v}_2^{(j)}k + \gamma_i^{(j)}\mathbf{v}_3^{(j)}(k) \quad (2)$$

where $\alpha_i^{(j)}, \beta_i^{(j)}$ and $\gamma_i^{(j)}$ are the *affine invariant* coordinates of $\mathbf{P}_i^{(j)}$ with respect to the basis $\mathbf{b}^{(j)}$.

Next, assume that the position of each point at time $k+1$ is related to its past positions $\mathbf{p}(k-i)$ by an ARMAX model

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

of the form:

$$\mathbf{p}(k+1) = \sum_{i=0}^{m-1} g_i \mathbf{p}(k-i) + h_i \mathbf{e}(k-i) \qquad (3)$$

where g_i, h_i are fixed coefficients and $\mathbf{e}(.)$ denotes a stochastic input. Note that this can be always assumed without loss of generality, since given N_F measurements of $\mathbf{p}(.), \mathbf{e}(.)$, there always exists a linear operator such that (3) is satisfied ([16], Chapter 10). Collecting all measurements in a vector $\mathbf{y}$ and taking z-transforms leads to a model of the form:

$$\mathbf{y} = \mathcal{M}\mathbf{e} \qquad (4)$$

where $\mathcal{M}$ is a linear time invariant operator that has a state space representation of the form:

$$\xi(k+1) = \mathbf{A}\xi(k) + \mathbf{B}\mathbf{e}(k), \qquad \mathbf{y}(k) = \mathbf{C}\xi(k) \qquad (5)$$

where $\mathbf{y}(k) = \begin{bmatrix} \mathbf{p}_1^{(1)}(k)^T & \mathbf{p}_2^{(1)}(k)^T & \dots & \mathbf{p}_n^{(m)}(k)^T \end{bmatrix}^T$ and $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ are suitable matrices. Moreover, by using a similarity transformation if necessary, the states ξ of the realization can be chosen so that they correspond to present and past values of the position of the centroids $\mathbf{O}^{(j)}$ and the vectors defining the bases $\mathbf{b}^{(j)}$. Finally, by absorbing if necessary the spectral density of $\mathbf{e}$ in $\mathcal{M}$, it can always be assumed that $\mathbf{e}(.)$ is an impulse. For example, in the simple case of constant angular velocity ω, the state–space representation associated with the two left–wing propeller tracks shown in Figure 1 is given by:

$$\begin{aligned}
\xi(k) &= \begin{bmatrix} \mathbf{v}^{(1)}(k) & \mathbf{v}^{(2)}(k) \end{bmatrix}^T \\
\mathbf{v}^{(i)}(k) &\doteq \begin{bmatrix} v_1^{(i)}(k) & v_2^{(i)}(k) & t_x^{(i)}(k) \end{bmatrix} \\
\mathbf{A} &= \begin{bmatrix} \mathbf{R} & 0 \\ 0 & \mathbf{R} \end{bmatrix}, \ \mathbf{R} = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
\mathbf{B} &= \xi(0), e(k) = \begin{cases} 1 \text{ if } k = 0 \\ 0 \text{ otherwise} \end{cases} \\
\mathbf{C} &= \begin{bmatrix} \alpha_1^{(1)} & \beta_1^{(1)} & 1 & 0 & 0 & 0 \\ \alpha_2^{(1)} & \beta_2^{(1)} & 1 & 0 & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \alpha_1^{(2)} & \beta_1^{(2)} & 1 \\ 0 & 0 & 0 & \alpha_2^{(2)} & \beta_2^{(2)} & 1 \\ \vdots & & & & & \vdots \end{bmatrix}
\end{aligned} \qquad (6)$$

Note that in this coordinate system the operator $\mathcal{M}$ has a block diagonal form (see Figure 2). This is due to the fact that, for any two points $\mathbf{p}_r^{(j)}, \mathbf{p}_s^{(j)}$ belonging to the same object, the time evolution of the *relative difference* $\delta_{r,s}(k) \doteq \mathbf{p}_r^{(j)}(k) - \mathbf{p}_s^{(j)}(k)$ can be explained *only* in terms of the trajectories of the basis $\mathbf{b}^{(j)}$. On the other hand, modelling the evolution of the difference between points belonging to two different objects requires considering $\mathbf{O}^{(i)}, \mathbf{O}^{(j)}, \mathbf{b}^{(i)}, \mathbf{b}^{(j)}$. Formally, all the states of $\mathbf{A}$ unrelated to $\mathbf{b}^{(j)}$ are *unobservable* from $\delta_{r,s}$, and thus the subsystem $\mathcal{M}_{r,s}\colon \mathbf{e} \to \delta_{r,s}$ is rank deficient when compared to a subsystem mapping $\mathbf{e}$ to the difference between points on different objects. For instance, in the propeller example used above, the subsystem mapping $\mathbf{e}$ to the difference between any two points $\mathbf{p}_i^{(1)}, \mathbf{p}_j^{(1)}$ on the *same* propeller has a state–space realization $(\mathbf{A}, \mathbf{B}, \mathbf{C}_{i,j})$, where $\mathbf{A}, \mathbf{B}$ are given in (6) and

$$\mathbf{C}_{i,j} \doteq \begin{bmatrix} \alpha_i^{(1)} - \alpha_j^{(1)} & \beta_i^{(1)} - \beta_j^{(1)} & 0 & 0 & 0 & 0 \end{bmatrix}$$

Note that the last 4 elements of $\mathbf{C}_{i,j}$ are zero, which, coupled with the block-diagonal structure of $\mathbf{A}$ implies that the last four states in ξ are *unobservable* from $\delta = \mathbf{p}_i^{(1)} - \mathbf{p}_j^{(1)}$. Hence, a minimal realization of $\mathcal{M}\colon e \to \delta$ is given by $\mathcal{M}_{red} \doteq (\mathbf{A}_{red}, \mathbf{B}_{red}, \mathbf{C}_{red})$, where $\mathbf{A}_{red} = \mathbf{R}$, $\mathbf{C}_{red} \doteq \begin{bmatrix} \alpha_i^{(1)} - \alpha_j^{(1)} & \beta_i^{(1)} - \beta_j^{(1)} \end{bmatrix}$ and $\mathbf{B}_{red} \in R^{2\times 1}$ is a matrix related to the initial conditions. This realization has rank 2[3], compared to the case of points on different propellers that has (generically) rank 6. Roughly speaking, the relative motion of points in a given object, carries no information about the motion of other objects. It follows that points can be assigned to objects by searching for minimum rank clusters of the operator $\mathcal{M}_{i,j}$.

3. Segmentation Algorithm

In this section we show that the dynamic rank based segmentation idea outlined above can be be accomplished in an efficient, robust way by computing the singular values of a sequence of matrices (Hankel matrices) constructed from the measured correspondence data.

3.1. Estimating Model Order from Experimental Data

Next we recall a result that allows for efficiently solving the problem of finding minimum rank clusters of the operator $\mathcal{M}$, based solely on the measured data.

Lemma 1. *[13]. Consider a system with a minimal state space realization of the form* (5) *with* $A \in R^{n\times n}$ *and* $B \in$

[3] Recall that the rank of $\mathcal{M}$ is given by the number of states of a minimal (observable and controllable) realization ([16], Chapter 9).

$R^{n\times m}$. *Form the infinite matrix* $\mathbf{H} = \begin{pmatrix}\mathbf{H_y}\\ \mathbf{H_e}\end{pmatrix}$ *where*

$$\mathbf{H_y} = \begin{bmatrix} \mathbf{y}(1) & \mathbf{y}(2) & \dots & \mathbf{y}(m) & \dots \\ \mathbf{y}(2) & \mathbf{y}(3) & \dots & \mathbf{y}(m+1) & \dots \\ \vdots & \dots & \vdots & \vdots & \\ \mathbf{y}(n) & \mathbf{y}(n+1) & \dots & \mathbf{y}(m+n) & \dots \\ \vdots & \dots & \vdots & \vdots & \end{bmatrix}$$

$$\mathbf{H_e} = \begin{bmatrix} \mathbf{e}(1) & \mathbf{e}(2) & \dots & \mathbf{e}(m) & \dots \\ \mathbf{e}(2) & \mathbf{e}(3) & \dots & \mathbf{e}(m+1) & \dots \\ \vdots & \dots & \vdots & \vdots & \\ \mathbf{e}(n) & \mathbf{e}(n+1) & \dots & \mathbf{e}(m+n) & \dots \\ \vdots & \dots & \vdots & \vdots & \end{bmatrix}; \quad (7)$$

Then if $\mathcal{E}\{\xi(k)\mathbf{e}(k)^T\} = 0$*, where* $\mathcal{E}$ *denotes expectation,* $rank(H) = n + m$*.*

Remark 1. *In the case under consideration here,* $\mathbf{e}(1) = 1, \mathbf{e}(k) = 0, k > 1$ *and thus rank*$(\mathbf{H}_y) = n$*. It follows that given* N_F *frames, assuming that* $N_F \gg n$ *the rank of the operator* $\mathcal{M}$ *can be estimated by computing the rank of the finite Hankel matrix:*

$$\mathbf{H_y} = \begin{bmatrix} \mathbf{y}(1) & \mathbf{y}(2) & \dots & \mathbf{y}(\frac{N_F}{2}) \\ \mathbf{y}(2) & \mathbf{y}(3) & \dots & \mathbf{y}(\frac{N_F}{2}+1) \\ \vdots & \dots & \dots & \vdots \\ \mathbf{y}(\frac{N_F}{2}) & \dots & \dots & \mathbf{y}(N_F) \end{bmatrix}$$

This observation allows for performing the dynamic motion segmentation proceeding as follows:

(i) Given a set of n points (possibly coming from more than one object) labelled across N_F frames, consider the pairwise differences $\delta_{r,s}(k) = \mathbf{p}_r(k) - \mathbf{p}_s(k)$ and compute the rank of the corresponding Hankel matrix $\mathbf{H}_{\delta_{r,s}}$.

(ii) Assign each point to the group corresponding to the minimum value of rank$[\mathbf{H}_\delta(i,j)]$.

Note that in principle, the algorithm above requires noiseless data, since rank computations are very sensitive to the presence of noise and outliers. As we show next, these effects can be robustly handled by simply replacing "rank" by the number of singular values below a certain threshold related to the covariance of the measurement noise. To show this, assume that the measurements are corrupted by additive, zero mean, noise $\eta(.)$, e.g., $\hat{y}(k) = y(k) + \eta(k)$. In this case, $\mathbf{H}_{\hat{y}} = \mathbf{H}_y + \mathbf{H}_\eta$, where $\mathbf{H}_\eta$ is the Hankel matrix associated with the noise sequence $\eta(.)$. Moreover, it can be easily shown that, under ergodicity assumptions, $\mathbf{H}_\eta^T\mathbf{H}_\eta$ is an estimate of the covariance matrix of the noise, Σ_η. It follows that noisy measurements can be handled by simply replacing rank$(\mathbf{H}_\delta)$ by n_i, the number of singular values $\sigma_i(\mathbf{H}_\delta) \geq \sigma_\eta$, where σ_η^2 is an upper bound of the variance of the measurement noise. Finally, outliers are easily detected by the proposed algorithms as points that are uncoupled (or weakly coupled) to every other point, that is points where the rank of the relative Hankel matrix remains high and relatively unchanged for every possible pair.

3.2. Hankel Matrix Based Motion Segmentation

Based on the considerations above, we propose the following motion-segmentation algorithm:

Algoritm: SEGMENTMOTION

Input. (i) $\mathcal{W}$: the measurements matrix, where $\mathbf{w}_t^i = \begin{bmatrix} u_t^i \\ v_t^i \end{bmatrix}$ is the i^{th} point position in the t^{th} frame.
N_p: number of features.
N_F: number of frames.
(ii) σ_n: noise standard deviation.
Output. $\mathbf{\Gamma}$: Sorted coupling matrix.

for all $i \neq j \in \{1, \cdots, N_p\}$ **do**

$$\mathbf{H} \leftarrow \begin{bmatrix} d_1 & d_2 & \cdots & d_{\frac{N_F}{2}} \\ d_2 & d_3 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ d_{\frac{N_F}{2}} & \cdots & \cdots & d_{N_F} \end{bmatrix}$$

where $d_t = \begin{bmatrix} \mathbf{w}_t^i - \mathbf{w}_t^j \end{bmatrix}$

Compute $\mathbf{H} = \mathbf{UDV}^T$ using SVD.

$\mathbf{\Gamma}_{ij} \leftarrow$ number of singular values $\geq \sigma_n$.

end for

reorder $\mathbf{\Gamma}$ **using the approach in [5]**

If necessary, additional robustness can be achieved, at the price of increased computational complexity, by considering triplets, rather than pairs, of points.

4. Experimental Results

In this section we illustrate the proposed method with several examples involving both degenerate and non–degenerate motion, and compare the results against those obtained using currently existing methods. In all cases except Example 1, the multi-frame point correspondences where estimated by the Lucas-Kanade optical flow algorithm [12], and the dynamic rank matrix obtained using our algorithm was sorted using the approach suggested in [5]. Darker regions of the resulting matrix correspond to low values of the rank of the corresponding Hankel matrix indicating groups of points that can be *jointly* explained by a simpler dynamic model. On the other hand, light regions correspond to high complexity models, indicating features

likely to belong to different objects.

Example 1: Simple Dependent Motions. Figure 3(a) shows the results of applying the proposed segmentation method to the airplane example discussed in section 1. As shown in the figure, the proposed method achieves perfect segmentation, even though the motion is degenerate. For comparison, methods that rely solely on factorizations of subspaces of $\mathcal{W}$, fail to correctly segment the objects.

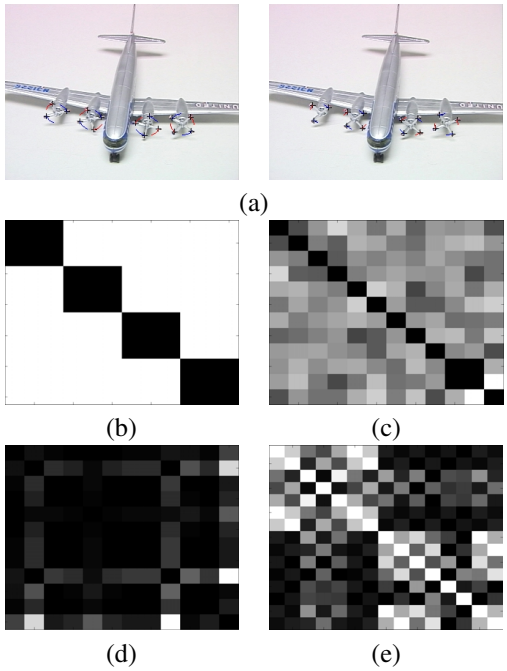

Figure 3. (a) All propellers move at the same speed. Right wing propellers move counterclockwise, while left wing propellers move clockwise. (b) Dynamics based segmentation. (c) Costeira-Kanade segmentation. (d) Zelnik-Manor-Irani segmentation using six eigenvectors. (e) GPCA segmentation similarity matrix.

Example 2: Dependent Motion in Cluttered Scenes. In this example we consider a more complex case: a person walking in a cluttered environment. Note that the motion is still degenerate, since different body parts share a common translation and some portion of the rotations. Figure 4(b) shows the results of applying the proposed algorithm. The segmented block on the upper left corner of the matrix corresponds to background features. These static features share the simplest possible dynamic model, with rank 1. The slightly lighter area around the background labelled region corresponds to features that almost share the background model for dynamics but need a slightly more complex model. In this scene they correspond to features that are improperly tracked due to interference from the foreground object, and can be safely discarded. This illustrates the ability of our method to exploit dynamical information to identify outliers, by pointing out features that have some association with some group but require higher complexity models than the rest of the features there. Note that this pattern also occurs within each of the labeled groups. Finally, features that do not show strong coupling with any other feature, i.e. columns of almost constant (high) order, can also be considered outliers and discarded. This is the case of some of the features between the first and second block in figure 4(b).

Example 3: Dependent and Independent Motions in Cluttered Scenes. In this case, the experimental data, shown in Figure 5, consists of a video sequence of a box with checkerboard pattern moving on a turntable, a ball moving with a complex translational/rotational motion and a cup moving with a translational motion on a cluttered environment. The features tracked in this sequence belong only to the three different objects tracked and the background, there are no outliers. The tracking for most of the features is accurate. Even though the distinct objects in this sequence present widely different motion models, they all share some component of their motion model. The box and the background features do not present translational motion, the background and the cup do not present rotational motion, and finally the ball presents a small translational motion with respect to the rotational motion. As expected the objects in this scene can not be segmented based on the structure of the shape interaction matrix. GPCA correctly identifies some of the objects, but incorrectly groups some background points together with the box and splits the ball into two clusters. On the other hand, the proposed dynamics–based approach yields the correct segmentation.

5. Conclusions

In this paper we propose a motion segmentation algorithm based upon the idea of grouping points according to the complexity of the model required to explain their relative motion. Intuitively, this formalizes the idea that points on the same rigid share more modes of motion (for instance a common translation or rotation) than points on different objects, leading to less complex models. By exploiting results from systems theory, the problem of robustly estimating the complexity of the underlying model is reduced to computing the rank of a Hankel matrix constructed from the correspondence data. Contrary to the case when working directly with the measurements matrix $\mathcal{W}$ whose rank is invariant under a column permutation, the rank of the Hankel matrix is a function of both *geometric* and *temporal* constraints. This ability to exploit both sets of constraints renders the proposed method less sensitive to the effect of noise or outliers than approaches that rely solely on factorizations of $\mathcal{W}$. In addition, the method can also naturally handle "degenerate cases", e.g. cases where the objects partially share motion modes. Note however that this additional robustness comes at the price of increased computational complexity, since the method requires computing N_p^2 SVDs (albeit of

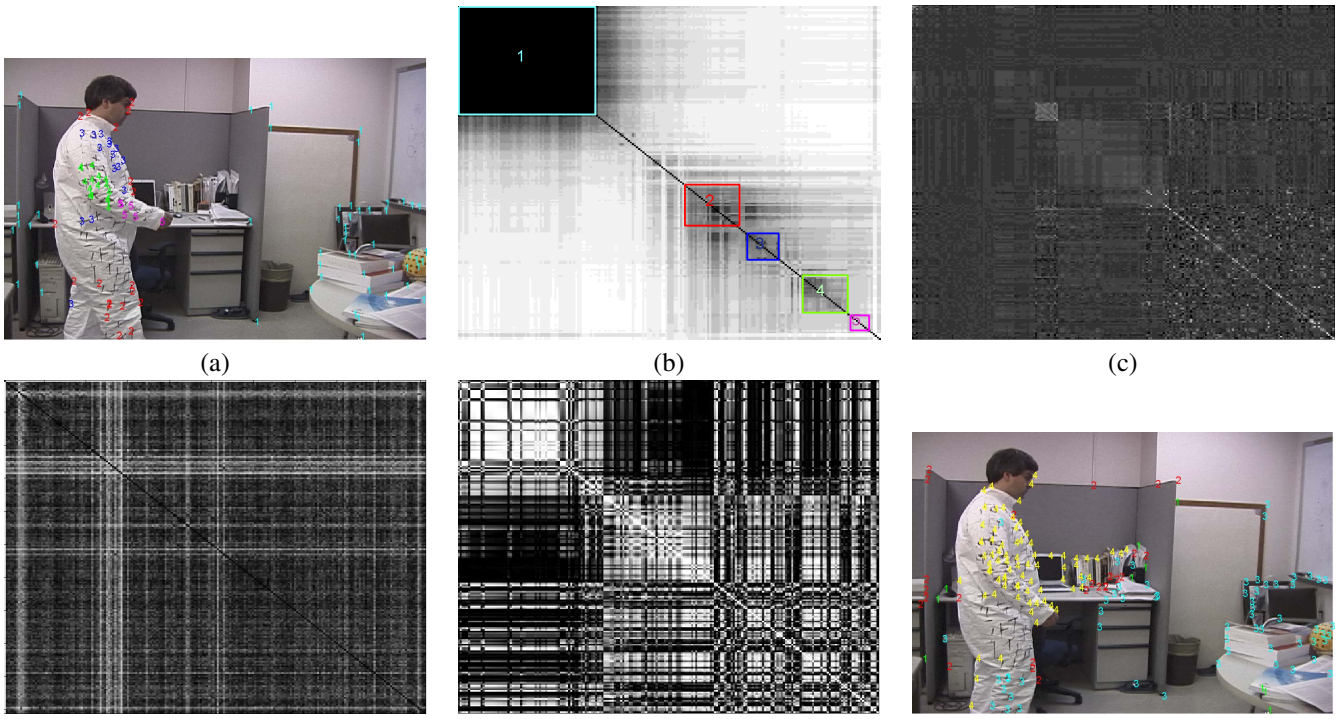

Figure 4. (a) A person walking in a cluttered environment. (b) Dynamics based segmentation. (c) Costeira-Kanade segmentation. (d) Zelnik-Manor-Irani segmentation using fifteen eigenvectors. (e) GPCA segmentation. (f) GPCA labelling.

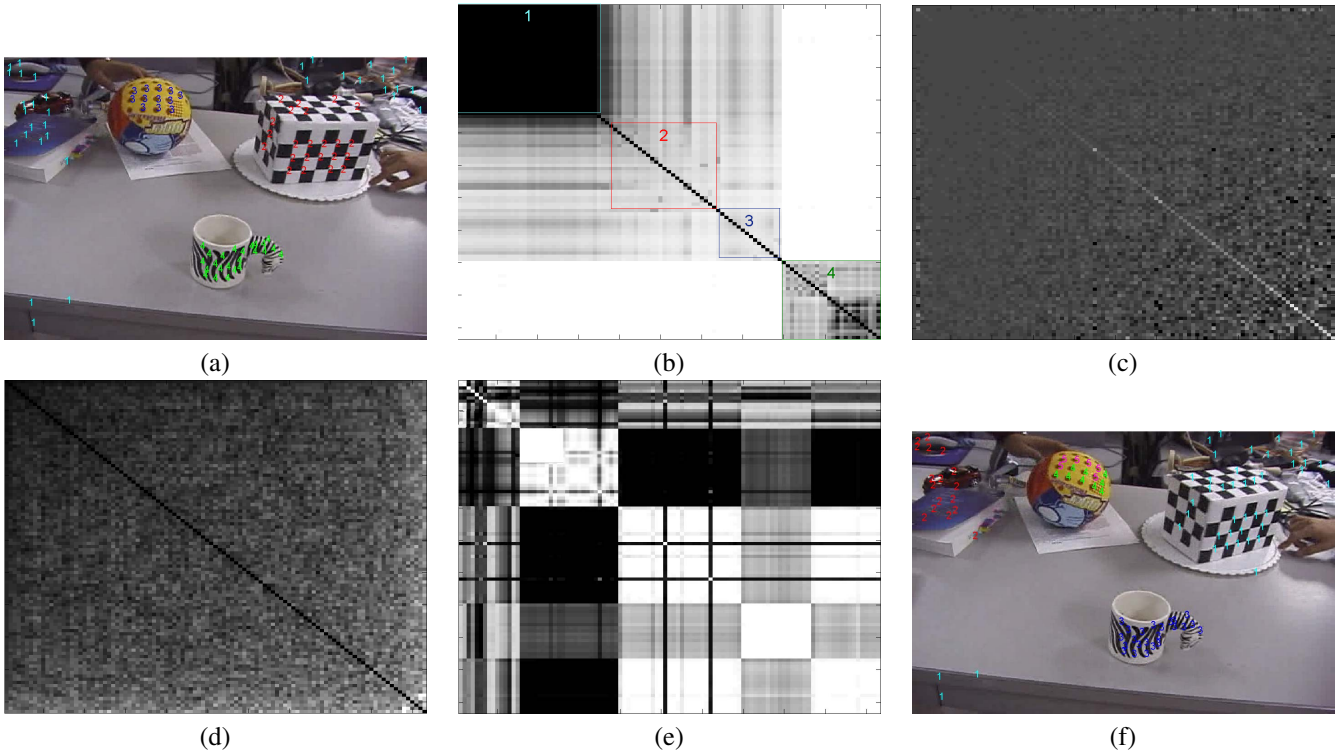

Figure 5. (a) Several objects moving with differently in a cluttered environment. (b) Dynamics based segmentation. (c) Costeira-Kanade segmentation. (d) Zelnik-Manor-Irani segmentation using fifteen eigenvectors. (e) GPCA segmentation similarity matrix. (f) GPCA labelling.

matrices considerably smaller than $\mathcal{W}$). If necessary, this computational complexity can be reduced by using a factorization based approach to perform an initial segmentation and then applying the proposed algorithm to each of the resulting clusters.

Research is currently underway to extend the proposed method to handle the case of missing correspondences.

References

[1] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):42–77, 1994.

[2] M. J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[3] M. J. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):972–986, October 1996.

[4] C. Cedras and M. Shah. Motion-based recognition: A survey. *Image Vis. Comput.*, 13:129–155, March 1995.

[5] J. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision*, 29(3):159–179, September 1998.

[6] C. Gear. Multibody grouping from motion images. *International Journal of Computer Vision*, 29(3):133–152, August 1998.

[7] G. Halevy and D. Weinshall. Motion of disturbances: Detection and tracking of multibody nonrigid motion. *Machine Vision Applications*, 11(3):122–137, 1999.

[8] M. Han and T. Kanade. Multiple motion scene reconstruction with uncalibrated cameras. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(7):884–894, July 2003.

[9] R. Hartley and R. Vidal. The multipbody trifocal tensor: Motion segmentation from three perspective views. In *IEEE Computer Vision and Pattern Recognition*, volume 1, pages 769–775, 2004.

[10] R. Jain, R. Kasturi, and B. G. Schunck. *Machine Vision*. Mac Graw Hill, 1995.

[11] K. Kanatani. Motion segmentation by subspace separation: Model selection and reliability evaluation. *International Journal of Image and Graphics*, 2(2):179–197, April 2002.

[12] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI81*, pages 674–679, 1981.

[13] M. Moonen, B. D. Moor, L. Vandenberghe, and J. Vandewalle. On- and off-line identification of linear state-space models. *International Journal of Control*, 49(1):219–232, 1989.

[14] T. Morita and T. Kanade. A paraperspective factorization method for recovering shape and motion from image sequences. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(8):858–867, August 1997.

[15] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(3):206–218, 1997.

[16] R. Sánchez Peña and M. Sznaier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.

[17] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[18] C. Stauffer and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

[19] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 709–720. Sprin-ger-Ver-lag, April 1996.

[20] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

[21] P. H. S. Torr. Geometric motion segmentation and model selection. *Phil. Trans. Royal Society of London*, 356(1740):1321–1340, 1998.

[22] L. Torresani, D. Yang, G. Alexander, and C. Bregler. Tracking and modelling non-rigid objects with rank constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 493–500, Kaui, Hawaii, December 2001.

[23] R. Vidal and R. Hartley. Motion segmentation with missing data using powerfactorization and gpca. In *IEEE Computer Vision and Pattern Recognition*, volume 2, pages 310–316, 2004.

[24] R. Vidal, Y. Ma, S. Soatto, and S. Sastry. Two–view multibody structure from motion. *International Journal of Computer Vision*, to appear, 2005.

[25] J. Y. A. Wang and E. H. Adelson. Representing Moving Images with Layers. *The IEEE Transactions on Image Processing Special Issue: Image Sequence Compression*, 3(5):625–638, September 1994.

[26] Y. Weiss. Segmentation using eigenvectors: a unifying view. In *IEEE International Conference on Computer Vision*, pages 975–982, 1999.

[27] Y. Wu, Z. Zhang, T. S. Huang, and J. Lin. Multibody grouping via orthogonal subspace decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume II, pages 252–257, Kaui, Hawaii, December 2001.

[28] J. Xiao, J. Chai, and T. Kanade. A closed-form solution to non-rigid shape and motion recovery. In *The 8th European Conference on Computer Vision (ECCV 2004)*, May 2004.

[29] L. Zelnik-Manor and M. Irani. Degeneracies, dependencies and their implications in multi-body and multi-sequence factorization. In *IEEE Computer Vision and Pattern Recognition*, pages 287–293, 2003.

[30] H. Zhou and T. S. Huang. Recovering articulated motion with a hierarchical factorization method. In 5^{th} *International Workshop on Gesture and Sign Language based Human–Computer Interaction*, Genoa, April 2003.