# A Factorization Method for 3D Multi-body Motion Estimation and Segmentation*

René Vidal
Department of EECS
University of California
Berkeley – CA 94720
rvidal@eecs.berkeley.edu

Stefano Soatto
Dept. of Computer Science
University of California
Los Angeles – CA 90095
soatto@ucla.edu

S. Shankar Sastry
Department of EECS
University of California
Berkeley – CA 94720
sastry@eecs.berkeley.edu

**Abstract**

We study the problem of estimating the motion of independently moving objects observed by a moving perspective camera. We show that infinitesimal image measurements corresponding to independent motions lie on orthogonal six-dimensional subspaces of a higher-dimensional linear space. We propose a factorization algorithm that estimates the number of independent motions, the segmentation of the image points and the motion of each object relative to the camera from a set of image points and their optical flows in multiple frames. We evaluate the proposed algorithm on synthetic and real image sequences.

## 1 Introduction

The problem of estimating the 3D motion of a moving camera imaging a single static object has been thoroughly studied in the computer vision community. One of the first multi-view algorithms was proposed by Tomasi and Kanade [15] who used a factorization technique based on the fact that, under orthographic projection, discrete image measurements lie on a three-dimensional linear variety. The method was extended to affine and paraperspective cameras in [11] and to central panoramic cameras in [12].

Factorization approaches have also been extended to the case of multiple moving points/objects observed by an *orthographic* camera. Boult and Brown [2] proposed a rank constraint to estimate the number of independent motions and obtained the segmentation of the image data from the leading singular vectors of the matrix of feature points in multiple frames. Their algorithm was extended by Costeira and Kanade [3] who showed that subspaces corresponding to different motions are orthogonal to each other. They used this fact to define the so-called *interaction matrix*, from which the segmentation of the image data is obtained using a graph-theoretic approach. A similar algorithm was proposed by Han and Kanade [5] for reconstructing a scene containing multiple moving points, some of them static and the others moving linearly with constant speed.

Under full perspective projection, discrete image measurements form a nonlinear variety [16]. Therefore, even though factorization methods have been used for single-body motion estimation [10], they have not been generalized to the case of multiple motions

*Research supported by ONR N00014-00-1-0621, ARO DAAD19-99-1-0139, NSF ECS-0200511.

yet. Instead, various special cases have been analyzed using geometric techniques, *e.g.* multiple points moving linearly with constant speed [5, 13] or in a conic section [1], and two-body [18] and multi-body [17] motion segmentation from two perspective views. Alternative probabilistic approaches to 3-D motion segmentation are based on model selection techniques [16, 7], combine normalized cuts with a mixture of probabilistic models [4], or compute statistics of the innovation process of a recursive filter [14].

On the other hand, perspective infinitesimal image measurements do lie on a low-dimensional linear variety defined by the so-called *subspace constraints* [6]. Irani used these constraints to obtain a multi-frame algorithm for the estimation of the optical flow of a moving camera observing a static scene. However, these constraints have not been used for 3D motion estimation.

In this paper, we use subspace constraints to develop a factorization algorithm for 3D motion estimation and segmentation from multiple perspective views. We do not assume prior segmentation of the points, nor do we restrict the motion of the objects to be linear or constant. Also, we do not assume previous knowledge of the number of independent motions. Our approach is based on the fact that infinitesimal image measurements corresponding to independent motions lie on orthogonal six-dimensional subspaces of a higher-dimensional linear space. Therefore, one can estimate the number independent motions, the segmentation of the image points, and the motion of each object relative to the camera from a set of image points and their optical flows. We present experimental results on synthetic and real image sequences.

## 1.1 Notation and Problem Statement

The **motion** of the camera and that of the objects is modeled as a rigid body motion in $\mathbb{R}^3$, *i.e.* as an element of the special Euclidean group $SE(3) = \{(R,T) \mid R \in SO(3), T \in \mathbb{R}^3\}$ and its Lie algebra $se(3) = \{(\widehat{\omega}, v) \mid \widehat{\omega} \in so(3), v \in \mathbb{R}^3\}$, where SO(3) and so(3) are the sets of rotation matrices and skew-symmetric matrices in $\mathbb{R}^{3\times 3}$, respectively[1].

The **image** $\mathbf{x} = [x, y, 1]^T \in \mathbb{R}^3$ of a point $q$ with coordinates $[q_1, q_2, q_3]^T \in \mathbb{R}^3$ (with respect to the camera frame), is assumed to satisfy the *perspective projection* equation:

$$\mathbf{x} = q/Z, \tag{1}$$

where $Z = q_3 > 0$ encodes the (unknown and positive) **depth** of the point $q$ with respect to its image $\mathbf{x}$. The **optical flow** $\mathbf{u}$ is defined as the velocity of $\mathbf{x}$ on the image plane, *i.e.*

$$[\mathbf{u}^T, 0]^T = \dot{\mathbf{x}}.$$

**Problem Statement:** Let $\mathbf{x}_j^i$ be the image of point $i$ in frame $j$, with $i = 1, \ldots, n$ and $j = 0, \ldots, m$, where $j = 0$ indicates the reference frame. Let $\{\mathbf{u}_j^i\}$ be the optical flow of point $\mathbf{x}_0^i$ between frames 0 and $j = 1, \ldots, m$. Given the images $\{\mathbf{x}_0^i\}$ and the flows $\{\mathbf{u}_j^i\}$, recover the number of moving objects, the object to which each point belongs to, the depth of the $n$ points and the motion of the objects relative to the camera.

To be consistent with the notation, we always use the superscript to enumerate the $n$ different points and/or the object to which the point belongs to. We omit the superscript when we refer to a generic single point and/or object. The subscript is always used to enumerate the $m$ different camera frames.

[1]The "hat" notation, $\widehat{(\cdot)}$, denotes the map from $\mathbb{R}^3$ to so(3), that transforms a three-dimensional vector $u$ into a $3 \times 3$ matrix $\widehat{u}$ such that $\widehat{u}v = u \times v \;\forall\, u, v \in \mathbb{R}^3$.

## 2 Single-Body Multi-View Geometry

Let us start with the simplest case in which the moving camera observes a single moving object. Let $(R_o(t), T_o(t)) \in SE(3)$ and $(R_c(t), T_c(t)) \in SE(3)$ be the pose of the object and that of the camera at time $t$ with respect to an inertial (fixed) reference frame. Let $q$ be a point located on the object with coordinates $[q_1,\ q_2,\ q_3]^T \in \mathbb{R}^3$ relative to the object frame. The coordinates of the same point relative to the inertial reference frame are: $q_o(t) = R_o(t)q + T_o(t)$ and the coordinates of $q$ relative to the camera frame are: $q_{oc}(t) = R_c^T(t)R_o(t)q + R_c^T(t)(T_o(t) - T_c(t))$. Differentiating this equation yields: $\dot{q}_{oc} = (\dot{R}_c^T R_o + R_c^T \dot{R}_o)q + \dot{R}_c^T(T_o - T_c) + R_c^T(\dot{T}_o - \dot{T}_c)$. Combining the previous two equations yields:

$$\dot{q}_{oc} \quad = \quad (\dot{R}_c^T R_c + R_c^T \dot{R}_o R_o^T R_c) q_{oc} + R_c^T(\dot{T}_o - \dot{T}_c - \dot{R}_o R_o^T (T_o - T_c)). \tag{2}$$

Since $\dot{R}R^T \in so(3)$, $\widehat{R^T\omega} = R^T\widehat{\omega}R$ and $\dot{R}^T R = -R^T \dot{R} R^T R$ [9], we may define the angular velocities $\omega_c,\ \omega_o \in \mathbb{R}^3$ by: $\widehat{\omega_o} = \dot{R}_o R_o^T$ and $\widehat{\omega_c} = \dot{R}_c R_c^T$. Combining the previous equation with (2) yields:

$$\dot{q}_{oc} = [R_c^T(\omega_o - \omega_c)] \times q_{oc} + R_c^T(\dot{T}_o - \dot{T}_c - \widehat{\omega_o}(T_o - T_c)) = \widehat{\omega} q_{oc} + v,$$

where $\omega$ and $v$ are the angular and translational velocities of the object relative to the camera. Under perspective projection, the optical flow $\mathbf{u}$ of point $q$ is then given by:

$$\mathbf{u} = \frac{d}{dt}\left(\frac{q_{oc}}{Z}\right) = \frac{1}{Z}\begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \end{bmatrix} \dot{q}_{oc} = \begin{bmatrix} -xy & 1+x^2 & -y & 1/Z & 0 & -x/Z \\ -(1+y^2) & xy & x & 0 & 1/Z & -y/Z \end{bmatrix} \begin{bmatrix} \omega \\ v \end{bmatrix}$$

where $q_{oc} = (X, Y, Z)^T$ and $(x, y, 1)^T = q_{oc}/Z$.

Given measurements for the optical flow $\mathbf{u}_j^i = (\mathrm{u}_j^i, \mathrm{v}_j^i)^T$ of point $\mathbf{x}_0^i = (x^i, y^i, 1)^T$, $i = 1, \ldots, n$, in frame $j = 1, \ldots, m$, define the matrix of *rotational flows* $\Psi$ and the matrix of *translational flows* $\Phi$ as:

$$\Psi = \begin{bmatrix} -\{xy\} & \{1+x^2\} & -\{y\} \\ -\{1+y^2\} & \{xy\} & \{x\} \end{bmatrix} \in \mathbb{R}^{2n\times 3} \quad \text{and} \quad \Phi = \begin{bmatrix} \{1/Z\} & 0 & -\{x/Z\} \\ 0 & \{1/Z\} & -\{y/Z\} \end{bmatrix} \in \mathbb{R}^{2n\times 3},$$

where (for example) $\{xy\}^T = [x^1y^1, \cdots, x^ny^n]$. Also let

$$U = \begin{bmatrix} \mathrm{u}_1^1 & \cdots & \mathrm{u}_m^1 \\ \vdots & & \vdots \\ \mathrm{u}_1^n & \cdots & \mathrm{u}_m^n \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} \mathrm{v}_1^1 & \cdots & \mathrm{v}_m^1 \\ \vdots & & \vdots \\ \mathrm{v}_1^n & \cdots & \mathrm{v}_m^n \end{bmatrix}.$$

Then, the *optical flow* matrix $W \in \mathbb{R}^{2n\times m}$ satisfies:

$$W = \begin{bmatrix} U \\ V \end{bmatrix} = [\Psi \ \ \Phi]_{2n\times 6} \begin{bmatrix} \omega_1 & \cdots & \omega_m \\ v_1 & \cdots & v_m \end{bmatrix}_{6\times m} = SM^T$$

where $\omega_j$ and $v_j$ are the velocities of the object relative to the camera in the $j^{th}$ frame. We call $S \in \mathbb{R}^{2n\times 6}$ the *structure* matrix and $M \in \mathbb{R}^{m\times 6}$ the *motion* matrix. We conclude that, for general translation and rotation, the optical flow matrix $W$ has rank 6. This rank-6 constraint is an extension of the rank-3 constraint proposed by Oliensis [10], and was first derived by Irani [6] who used it to obtain a multi-frame algorithm for the estimation of the optical flow of a moving camera observing a static scene.

The rank constraint $\text{rank}(W) = 6$ can be naturally used to derive a factorization method for estimating the relative velocities $(\omega_j, v_j)$ and depth $Z^i$ from image points $\mathbf{x}_0^i$ and optical flows $\mathbf{u}_j^i$. We can do so by factorizing $W$ into its motion and structure components. To this end, consider the singular value decomposition (SVD) of $W = \mathcal{U}\mathcal{S}\mathcal{V}^T$ and let $\tilde{S} = \mathcal{U}$ and $\tilde{M} = \mathcal{V}\mathcal{S}$. Then we have $S = \tilde{S}A$ and $M = \tilde{M}A^{-T}$ for some $A \in \mathbb{R}^{6\times 6}$. Let $A_k$ be the $k$-th column of $A$. Then the columns of $A$ must satisfy:

$$\tilde{S}A_{1-3} = \Psi \quad \text{and} \quad \tilde{S}A_{4-6} = \Phi.$$

Since $\Psi$ is known, $A_{1-3}$ can be immediately computed. The remaining columns of $A$ and the vector of depths $\{1/Z\}$ can be obtained up to scale from:

$$\begin{bmatrix} -I & \tilde{S}_{\mathrm{u}} & 0 & 0 \\ -I & 0 & \tilde{S}_{\mathrm{v}} & 0 \\ \text{diag}(\{x\}) & 0 & 0 & \tilde{S}_{\mathrm{u}} \\ \text{diag}(\{y\}) & 0 & 0 & \tilde{S}_{\mathrm{v}} \\ 0 & \tilde{S}_{\mathrm{v}} & 0 & 0 \\ 0 & 0 & \tilde{S}_{\mathrm{u}} & 0 \end{bmatrix} \begin{bmatrix} \{1/Z\} \\ A_4 \\ A_5 \\ A_6 \end{bmatrix} = 0.$$

where $\tilde{S}_{\mathrm{u}} \in \mathbb{R}^{n\times 6}$ and $\tilde{S}_{\mathrm{v}} \in \mathbb{R}^{n\times 6}$ are the upper and lower part of $\tilde{S}$, respectively.

# 3 Multi-Body Multi-View Geometry

So far, we have assumed that the scene contains a single moving object. Now, we consider the case in which a single camera observes $n_o$ objects. The new optical flow matrix $W$ will contain additional rows corresponding to measurements from the different objects. However, we cannot directly apply the factorization method of the previous section to solve for the relative motion of each object, because we do not know which measurements in $W$ correspond to which object. We therefore need to consider the segmentation problem first, *i.e.* the problem of separating all the measurements into $n_o$ classes: $\mathcal{I}^k = \{i \in \{1...n\} | \ \mathbf{x}_0^i \in \text{object } k\}$. Furthermore, we assume that $n_o$ itself is unknown.

## 3.1 Estimating the number of independent motions

Assume that the camera tracks $n^k$ image points for object $k$ and let $n = \sum n^k$ be the total number of points tracked. Also let $U^k$ and $V^k$ be matrices containing the optical flow of object $k$. If the segmentation of these points were known, then the multi-body optical flow matrix could be written as:

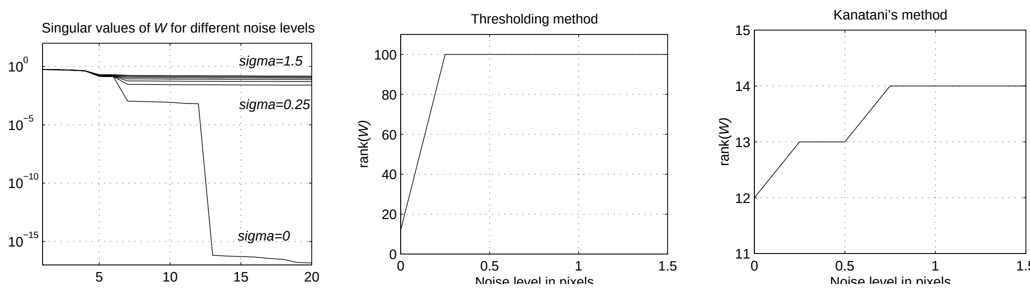
$$\begin{aligned} W &= \left[\begin{array}{c} U \\ \hline V \end{array}\right] = \left[\begin{array}{c} U^1 \\ \vdots \\ U^{n_o} \\ \hline V^1 \\ \vdots \\ V^{n_o} \end{array}\right] = \left[\begin{array}{ccc} \tilde{S}_{\mathrm{u}}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{S}_{\mathrm{u}}^{n_o} \\ \hline \tilde{S}_{\mathrm{v}}^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{S}_{\mathrm{v}}^{n_o} \end{array}\right] \begin{bmatrix} \tilde{M}^1 & \cdots & \tilde{M}^{n_o} \end{bmatrix}^T = \tilde{S}\tilde{M}^T \\ &= \tilde{S} \begin{bmatrix} A^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A^{n_o} \end{bmatrix} \begin{bmatrix} A^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A^{n_o} \end{bmatrix}^{-1} \tilde{M}^T = \tilde{S}AA^{-1}\tilde{M}^T = SM^T. \end{aligned}$$

where $\tilde{S}_{\mathrm{u}}^k$ and $\tilde{S}_{\mathrm{v}}^k \in \mathbb{R}^{n^k \times 6}$, $k = 1, \ldots, n_o$, $\tilde{S}$ and $S \in \mathbb{R}^{2n \times 6n_o}$, $A \in \mathbb{R}^{6n_o \times 6n_o}$, and $\tilde{M}$ and $M \in \mathbb{R}^{m \times 6n_o}$.

Since we are assuming that the segmentation of the image points is unknown, the rows of $W$ may be in a different order. However, the reordering of the rows of $W$ will not affect its rank. Assuming that $n \geq 6n_o$ and $m \geq 6n_o$, we conclude that the number of independent motions $n_o$ can be estimated as:

$$\boxed{n_o = \operatorname{rank}(W)/6.} \tag{3}$$

In practice, optical flow measurements will be noisy and $W$ will be full rank. Although one could estimate the number of objects by thresholding the singular values of $W$, a better choice comes from analyzing the statistics of the residual. Kanatani [8] studied the problem for the orthographic projection model using the geometric information criterion. The same method can be used here for a perspective camera as shown in Figure 1, which plots the singular values of $W$ and the estimated rank as a function of noise.




(a) (b) (c)

Figure 1: Estimating the rank of $W$ for two independent motions. Zero mean Gaussian noise with standard deviation $\sigma$ in pixels is added to $W$. (a) Singular values of $W$ for different levels of noise $\sigma \in [0, 1.5]$. (b) $\operatorname{rank}(W)$ estimated with a threshold of $10^{-4}$ and (c) with Kanatani's method.

## 3.2 Segmenting the image points

Segmenting the image points is equivalent to finding the unknown reordering of the rows of $W$. We can model such a reordering as an $n \times n$ permutation matrix $P$ applied to both $U$ and $V$. Such a permutation will affect the rows of $\tilde{S}$, hence those of $S$, but $A$, $\tilde{M}$ and $M$ are unaffected. Therefore, from the SVD of $W = \mathcal{U}\mathcal{S}\mathcal{V}^T$ we have

$$\mathcal{U}\mathcal{U}^T = \begin{bmatrix} P \begin{bmatrix} \tilde{S}_{\mathrm{u}}^1 \tilde{S}_{\mathrm{u}}^{1T} & & 0 \\ & \ddots & \\ 0 & & \tilde{S}_{\mathrm{u}}^{n_o} \tilde{S}_{\mathrm{u}}^{n_o T} \end{bmatrix} P^T & P \begin{bmatrix} \tilde{S}_{\mathrm{u}}^1 \tilde{S}_{\mathrm{v}}^{1T} & & 0 \\ & \ddots & \\ 0 & & \tilde{S}_{\mathrm{u}}^{n_o} \tilde{S}_{\mathrm{v}}^{n_o T} \end{bmatrix} P^T \\ P \begin{bmatrix} \tilde{S}_{\mathrm{v}}^1 \tilde{S}_{\mathrm{u}}^{1T} & & 0 \\ & \ddots & \\ 0 & & \tilde{S}_{\mathrm{v}}^{n_o} \tilde{S}_{\mathrm{u}}^{n_o T} \end{bmatrix} P^T & P \begin{bmatrix} \tilde{S}_{\mathrm{v}}^1 \tilde{S}_{\mathrm{v}}^{1T} & & 0 \\ & \ddots & \\ 0 & & \tilde{S}_{\mathrm{v}}^{n_o} \tilde{S}_{\mathrm{v}}^{n_o T} \end{bmatrix} P^T \end{bmatrix},$$

We define the *segmentation matrix* $\Sigma$ as the sum of the diagonal blocks of $\mathcal{U}\mathcal{U}^T$, *i.e.*

$$\Sigma = P \begin{bmatrix} \tilde{S}_{\mathrm{u}}^1 \tilde{S}_{\mathrm{u}}^{1T} + \tilde{S}_{\mathrm{v}}^1 \tilde{S}_{\mathrm{v}}^{1T} & & 0 \\ & \ddots & \\ 0 & & \tilde{S}_{\mathrm{u}}^{n_o} \tilde{S}_{\mathrm{u}}^{n_o T} + \tilde{S}_{\mathrm{v}}^{n_o} \tilde{S}_{\mathrm{v}}^{n_o T} \end{bmatrix} P^T.$$

Then, $\Sigma_{ij} > 0$ if and only if image points $i$ and $j$ belong to the same object. In the absence of noise, the matrix $\Sigma$ can be trivially used to determine the class to which each image point belongs to. One can also use each one of the two diagonal blocks of $\mathcal{U}\mathcal{U}^T$. In the presence of noise, $\Sigma_{ij}$ will be nonzero even if points $i$ and $j$ correspond to different objects. Techniques that handle this case can be found in [3, 7] for the orthographic case. They can also be applied here to the perspective case.

# 4 Experimental Results

In this section, we evaluate the proposed algorithm on real and synthetic image sequences. Each pixel of each frame is considered as a feature and segmentation is performed using the segmentation matrix associated to the optical flow of those pixels.

Figure 2 shows the *street* sequence[2], which contains two independent motions: the motion of the car and the motion of the camera that is panning to the right. Figure 4(a) shows frames 3, 8 12 and 16 of the sequence with the corresponding optical flow superimposed. Optical flow is computed using Black's algorithm[3]. Figures 4(b)-(c) show the segmentation results. In frame 4 the car is partially occluded, thus only the frontal part of the car is segmented from the background. The door is incorrectly segmented because it is in a region with low texture. As time proceeds, motion information is integrated over time by incorporating optical flow from many frames in the optical flow matrix, thus the door is correctly segmented. In frame 16 the car is fully visible and correctly segmented from the moving background.

Figure 3 shows the *sphere-cube* sequence[2], which contains a sphere rotating along a vertical axis and translating to the right, a cube rotating counter clock-wise and translating to the left, and a static background. Even though the optical flow of the sphere appears to be noisy, its motion is correctly segmented. The top left (when visible), top and right sides of the square are also correctly segmented in spite of the fact that only normal flow is available. The left bottom side of the cube is merged with the background, because its optical flow is small, since the translational motion of the cube cancels its rotational motion. The center of the cube is never segmented correctly since it corresponds to a region with low texture. Integrating motion information over many frames does not help here since those pixels are in a region with low texture during the whole sequence.

Figure 4(a) shows the *two-robot* sequence with the corresponding optical flow superimposed. Figures 4(b) and 4(c) show the results of the segmentation. Groups 1 and 2 correspond to the each one of the moving objects, while group 3 corresponds to the background, which is the correct segmentation.

# 5 Conclusions

We have proposed an algorithm for estimating the motion of multiple moving objects as observed by a moving camera in multiple frames. Our algorithm is based on the fact that image measurements from independent motions lie on orthogonal subspaces of a higher-dimensional space, thus it does not require prior segmentation or previous knowledge of the number of independent motions. Experimental results show how segmentation is correctly obtained by integrating image measurements from multiple frames.

[2] http://www.cs.otago.ac.nz/research/vision/Research/OpticalFlow/opticalflow.html#Sequences
[3] http://www.cs.brown.edu/people/black/ignc.html

# References


[1] S. Avidan and A. Shashua. Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):348–357, 2000.

[2] T.E. Boult and L.G. Brown. Factorization-based segmentation of motions. In *Proc. of the IEEE Workshop on Motion Understanding*, pages 179–186, 1991.

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

[4] X. Feng and P. Perona. Scene segmentation from 3D motion. In *International Conference on Computer Vision and Pattern Recognition*, pages 225–231, 1998.

[5] M. Han and T. Kanade. Reconstruction of a scene with multiple linearly moving objects. In *International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 542–549, 2000.

[6] M. Irani. Multi-frame optical flow estimation using subspace constraints. In *IEEE International Conference on Computer Vision*, pages 626–633, 1999.

[7] K. Kanatani. Motion segmentation by subspace separation and model selection. In *IEEE International Conference on Computer Vision*, vol. 2, pages 586–591, 2001.

[8] K. Kanatani and C. Matsunaga. Estimating the number of independent motions for multibody motion segmentation. In *Asian Conference on Computer Vision*, 2002.

[9] R. M. Murray, Z. Li, and S. S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC press Inc., 1994.

[10] J. Oliensis. A multi-frame structure-from-motion algorithm under perspective projection. *International Journal of Computer Vision*, 34(2-3):163–192, 1999.

[11] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 19(3):206–18, 1997.

[12] O. Shakernia, R. Vidal, and S. Sastry. Infinitesimal motion estimation from multiple central panoramic views. In *Workshop on Motion and Video Computing*, 2002.

[13] A. Shashua and A. Levin. Multi-frame infinitesimal motion model for the reconstruction of (dynamic) scenes with multiple linearly moving objects. In *IEEE International Conference on Computer Vision*, volume 2, pages 592–599, 2001.

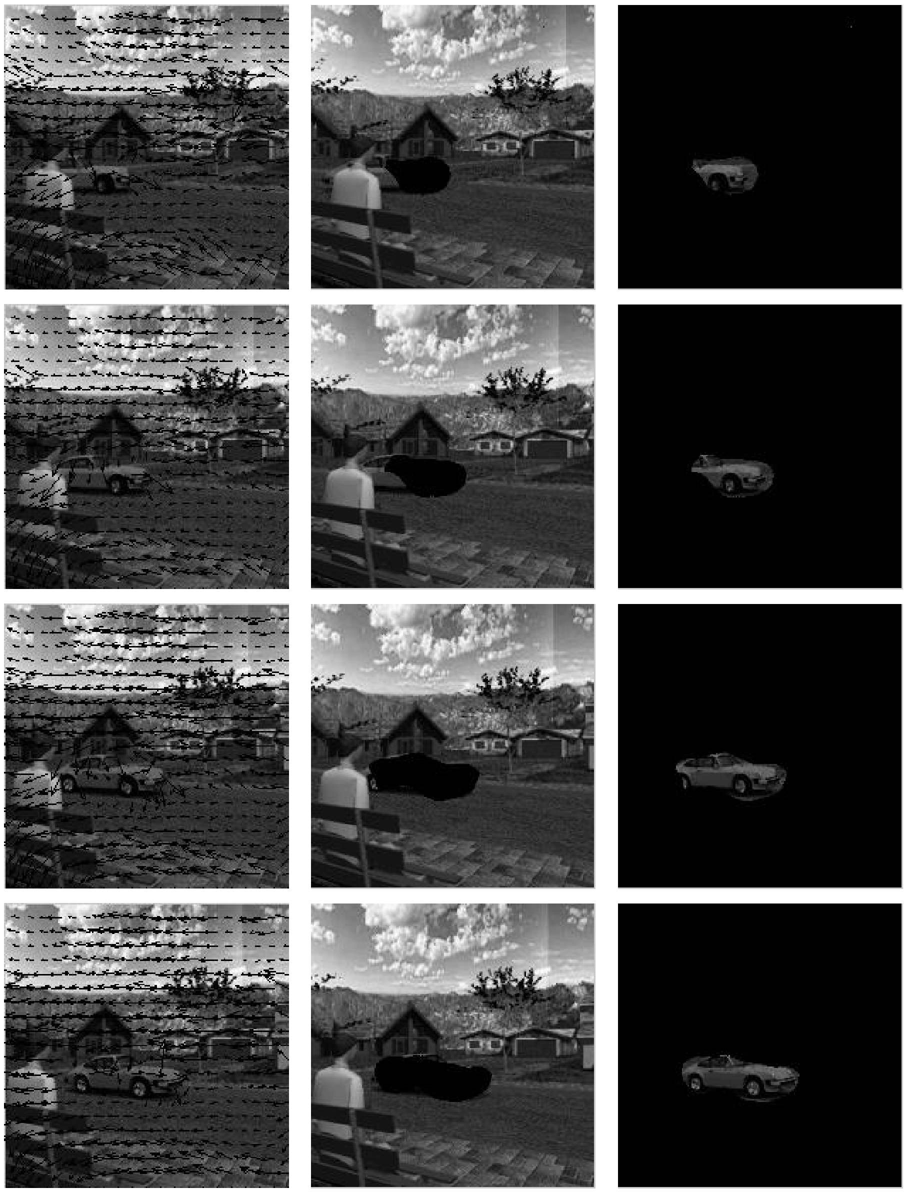
[14] S. Soatto and P. Perona. Three dimensional transparent structure segmentation and multiple 3D motion estimation from monocular perspective image sequences. In *IEEE Workshop on Motion of Nonrigid and Articulated Objects*, pages 228–235, 1994.

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

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

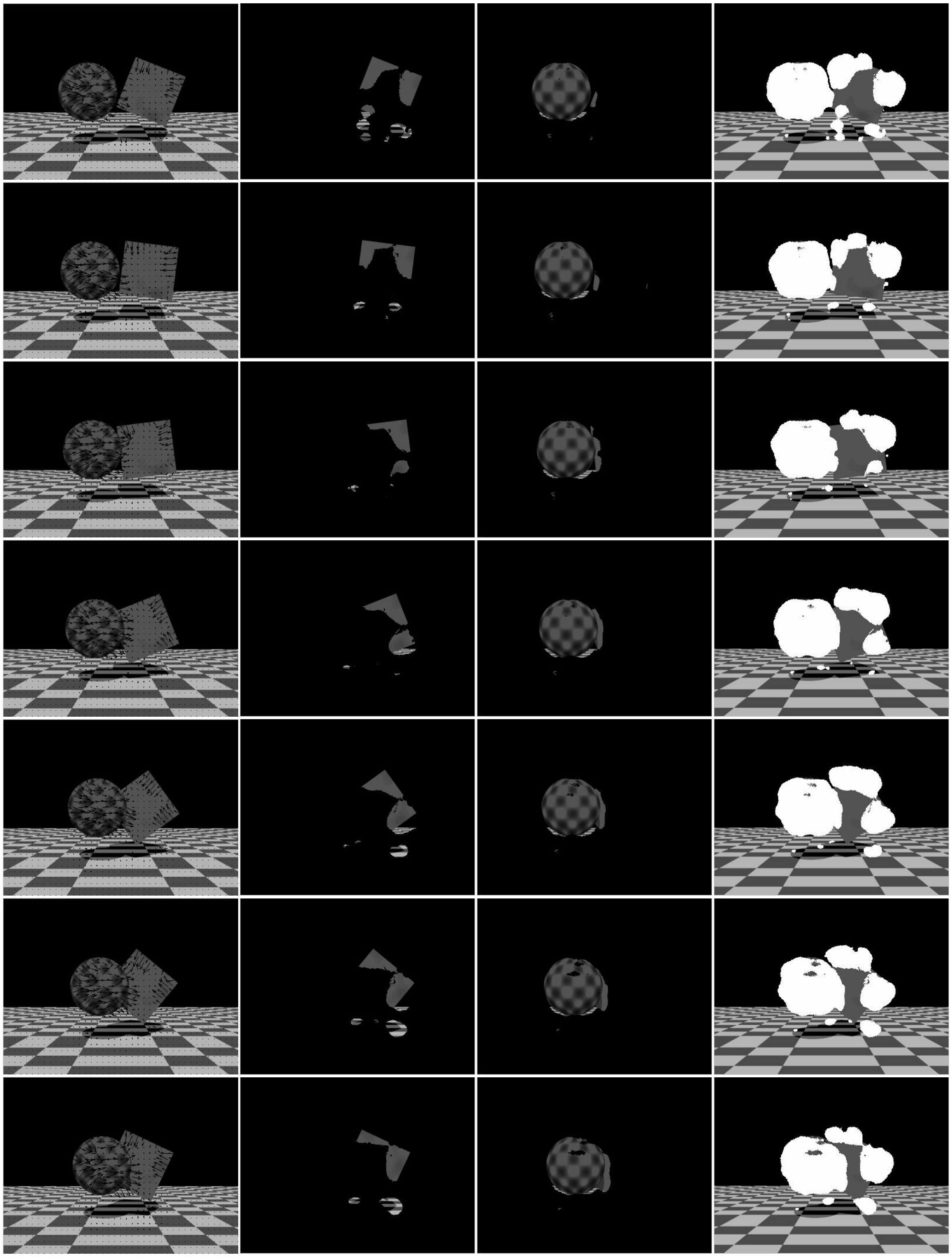
[17] R. Vidal, S. Soatto, Y. Ma, and S. Sastry. Segmentation of dynamic scenes from the multibody fundamental matrix. In *Workshop on Visual Modeling of Dynamic Scenes*, 2002.

[18] L. Wolf and A. Shashua. Two-body segmentation from two perspective views. In *International Conference on Computer Vision and Pattern Recognition*, 2001.



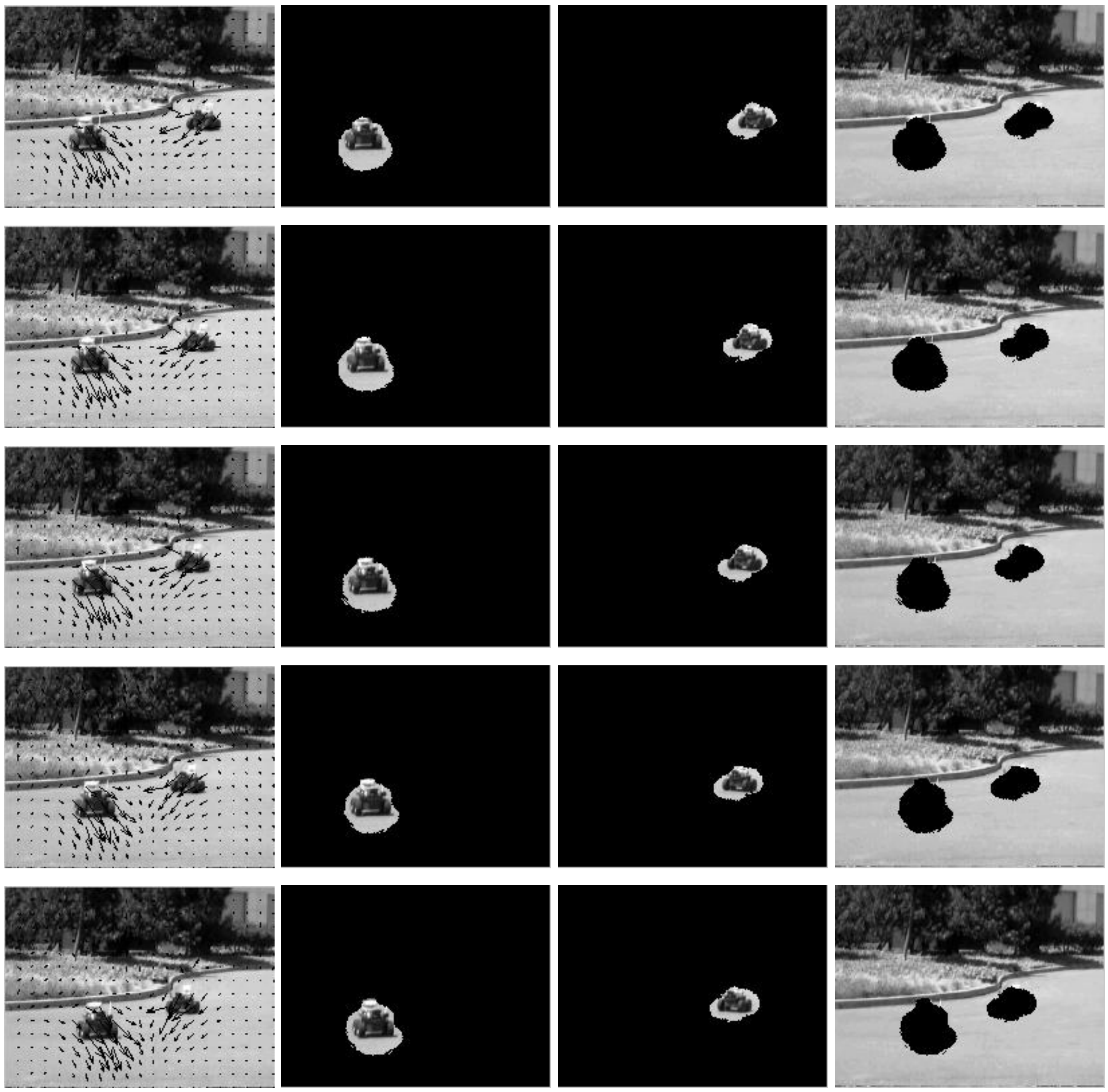
(a) Optical flow (b) Group 1 (c) Group 2

Figure 2: Segmentation results for the *street* sequence. The sequence has 18 frames and $200 \times 200$ pixels. The camera is panning to the right while the car is also moving to the right. (a) Frames 3, 8 12 and 16 of the sequence with the corresponding optical flow superimposed. (b) Group 1: motion of the camera. (c) Group 2: motion of the car.



(a) Optical flow (b) Group 1 (a) Group 2 (b) Group 3

Figure 3: Segmentation results for the *sphere-cube* sequence. The sequence contains 10 frames and $400 \times 300$ pixels. The sphere is rotating along a vertical axis and translating to the right. The cube is rotating counter clock-wise and translating to the left. The background is static. (a) Frames 2-8 with corresponding optical flow superimposed. (b) Group 1: cube motion. (c) Group 2: sphere motion. (d) Group 3: static background.



(a) Optical flow (b) Group 1 (c) Group 2 (d) Group 3

Figure 4: Segmentation results for the *two-robot* sequence. The sequence contains 6 frames and $200 \times 150$ pixels. (a) Frames 1-5 of the sequence with optical flow superimposed. (b) Group 1: one moving robot. (c) Group 2: the other moving robot. (d): Group 3: static background.