Multibody Motion Estimation and Segmentation from Multiple Central Panoramic Views

Omid Shakernia René Vidal Shankar Sastry

Department of Electrical Engineering & Computer Sciences
University of California at Berkeley, Berkeley CA 94720-1772
{*omids,rvidal,sastry*}*@eecs.berkeley.edu*

***Abstract*— We present an algorithm for infinitesimal motion estimation and segmentation from multiple central panoramic views. We first show that the central panoramic optical flows corresponding to independent motions lie in orthogonal ten-dimensional subspaces of a higher-dimensional linear space. We then propose a factorization-based technique that estimates the number of independent motions, the segmentation of the image measurements and the motion of each object relative to the camera from a set of image points and their optical flows in multiple frames. Finally, we present experimental results on motion estimation and segmentation for a real image sequence with two independently moving mobile robots, and evaluate the performance of our algorithm by comparing the vision estimates with GPS measurements gathered by the mobile robots.**

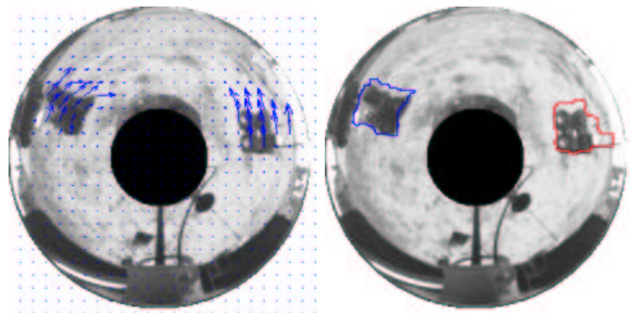

Fig. 1. Motion segmentation for two mobile robots based on their omnidirectional optical flows.

I. Introduction

The panoramic field of view offered by omnidirectional cameras makes them ideal candidates for many vision-based mobile robot applications, such as autonomous navigation, localization, and formation control. A problem that is fundamental to most of these applications is multibody motion estimation and segmentation, which is the problem of estimating of the number of independently moving objects in the scene; the segmentation of the objects from the background; and the relative motion between the camera and each one of the objects in the scene.

Most of the previous work in omnidirectional vision (see Section I-B) has been concerned with the special case in which a static scene is observed by a moving camera, the so-called *structure from motion problem*. The case in which both the camera and multiple objects move is more challenging, because one needs to simultaneously estimate multiple motion models without knowing which image measurements correspond to which moving object. Recent work has considered motion estimation and segmentation for orthographic and perspective cameras. To the best of our knowledge our work is the first one to address this problem in the case of central panoramic cameras.

A. Contributions of this paper

In this paper, we present the first algorithm for motion estimation and segmentation from multiple central panoramic views. Our algorithm estimates the number of independent motions, the segmentation of the image data and the motion of each object relative to the camera from a set of image points and their central panoramic optical flows in multiple frames. Our algorithm is based on a rank constraint on the optical flows generated by independently moving objects, which must lie in orthogonal ten-dimensional subspaces of a higher-dimensional linear space. We present experimental results on motion estimation and segmentation for a real image sequence with two independently moving mobile robots (see Fig. 1), and evalute the performance of our algorithm by comparing the vision estimates with differential GPS measurements gathered by the mobile robots. In [17], we applied the results of this work to vision-based formation control of nonholonomic mobile robots.

B. Previous Work

Central panoramic cameras are realizations of omnidirectional vision systems that combine a mirror and a lens and have a unique effective focal point. In [1], an entire class of omnidirectional cameras containing a single effective focal point is derived. A single effective focal point is necessary for the existence of epipolar geometry that is independent of the scene structure [11], making it possible to extend many results from perspective projection to the omnidirectional case.

The problem of estimating the 3D motion of a moving central panoramic camera imaging a single static object has received a lot of attention over the past few years. Researchers have generalized many two-view structure from motion algorithms from perspective projection to central panoramic projection, both in the case of discrete [5] and differential motion [7], [13]. In [7], [13], the image velocity vectors are mapped to a sphere using the Jacobian of the transformation between the projection model of the camera and spherical projection. Once the image velocities are on the sphere, one can apply well-known ego-motion algorithms for spherical projection. In [10], we proposed the first algorithm for motion estimation from *multiple* central panoramic views. Our algorithm does not need to map the image data onto the sphere, and is based on a rank constraint on the central panoramic optical flows which naturally generalizes the well-known rank constraints for orthographic [12], and affine and paraperspective [9] cameras.

The problem of estimating the 3D motion of multiple moving objects observed by a moving camera is, on the other hand, only partially understood. Costeira and Kanade [3] proposed a factorization method based on the fact that, under orthographic projection, discrete image measurements lie in a low-dimensional linear variety. Recently, Vidal *et al.* [18] (see also [8]) proposed a factorization algorithm for infinitesimal motion and perspective cameras based on the fact that independent motions lie in orthogonal subspaces of a higher dimensional space. The case of discrete image measurements was recently addressed by Vidal *et al.* [15], who proposed a linear algebraic algorithm based on the so-called *multibody epipolar constraint* between two perspective views of a scene. To the best of our knowledge, there is no work on motion segmentation from two or more central panoramic views, neither in the discrete not in the continuous case.

Paper Outline: In Section II we describe the projection model for central panoramic cameras and derive the optical flow equations. In Section III we present an algorithm for motion estimation and segmentation of multiple independently moving objects from multiple central panoramic views of a scene. In Section IV we specialize the algorithm to the case of planar motion in the $X-Y$ plane. In Section V we present experimental results evaluating the performance of the algorithm, and we conclude in Section VI.

II. Central Panoramic Imaging

In this section, we describe the projection model for a central panoramic camera and derive the central panoramic optical flow equations (see [10] for further details).

A. Projection Model

Catadioptric cameras are realizations of omnidirectional vision systems which combine a curved mirror and a lens. Examples of catadioptric cameras are a parabolic mirror in front of an orthographic lens and a hyperbolic mirror in front of a perspective lens. Camera systems with a unique effective focal point are called *central panoramic cameras*.

It was shown in [4] that all central panoramic cameras can be modeled by a mapping of a 3D point onto a sphere followed by a projection onto the image plane from a point in the optical axis of the camera. By varying two parameters (ξ, m), one can model all catadioptric cameras that have a single effective viewpoint. The particular values of (ξ, m) in terms of the shape parameters of different types of mirrors are listed in [2].

According to the unified projection model [4], the image point $(x, y)^T$ of a 3D point $q = (X, Y, Z)^T$ obtained through a central panoramic camera with parameters (ξ, m) is given by:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{\xi + m}{-Z + \xi\sqrt{X^2 + Y^2 + Z^2}} \begin{bmatrix} s_x X \\ s_y Y \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \quad (1)$$

where $0 \leq \xi \leq 1$, m, and (s_x, s_y) are scales that depend on the geometry of the mirror, the focal length and the aspect ratio of the lens, and $(c_x, c_y)^T$ is the mirror center.

Since central panoramic cameras for $\xi \neq 0$ can be easily calibrated from a single image of three lines [6], [2], in this paper, we assume that the camera has been calibrated, *i.e.* we know the parameters $(s_x, s_y, c_x, c_y, \xi, m)$. Therefore, without loss of generality, we consider the following *calibrated* central panoramic projection model:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} X \\ Y \end{bmatrix}, \quad \lambda \triangleq -Z + \xi\sqrt{X^2 + Y^2 + Z^2} \quad (2)$$

which is valid for $Z < 0$. It is direct to check that $\xi = 0$ corresponds to perspective projection, and $\xi = 1$ corresponds to paracatadioptric projection (a parabolic mirror in front of an orthographic lens).

B. Back-projection Rays

Since central panoramic cameras have a unique effective focal point, one can efficiently compute the *back-projection ray* (a ray from the optical center in the direction of the 3D point being imaged) associated with each image point.

One may consider the central panoramic projection model in equation (2) as a simple projection onto a curved virtual retina whose shape depends on the parameter ξ. We thus define the *back-projection ray* as the *lifting* of the image point $(x, y)^T$ onto this retina. That is, as shown in Fig. 2, given an image $(x, y)^T$ of a 3D point $q = (X, Y, Z)^T$, we define the back-projection rays as:

$$\mathbf{b} \triangleq (x, y, z)^T, \quad (3)$$

where $z = \mathbf{f}_\xi(x, y)$ is the height of the virtual retina. We construct $\mathbf{f}_\xi(x, y)$ in order to re-write the central panoramic projection model in (2) as a simple scaling:

$$\lambda\mathbf{b} = q, \quad (4)$$

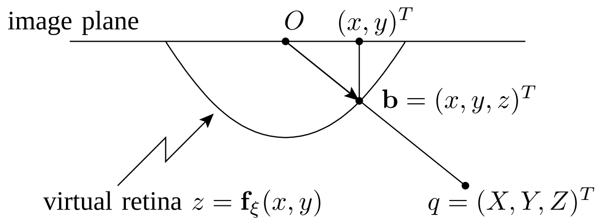

Fig. 2. Showing the curved virtual retina in central panoramic projection and back-projection ray $\mathbf{b}$ associated with image point $(x,y)^T$.

where the unknown scale λ is lost in the projection. Using equations (4) and (2), it is direct to solve for the height of the virtual retina as:

$$z \triangleq \mathbf{f}_\xi(x,y) = \frac{-1+\xi^2(x^2+y^2)}{1+\xi\sqrt{1+(1-\xi^2)(x^2+y^2)}}. \qquad (5)$$

Notice that in the case of paracatadioptric projection $\xi = 1$ and the virtual retina is the parabola $z = \frac{1}{2}(x^2+y^2-1)$.

C. Central Panoramic Optical Flow

If the camera undergoes a linear velocity $v \in \mathbb{R}^3$ and an angular velocity $\omega \in \mathbb{R}^3$, then the coordinates of a static 3D point $q \in \mathbb{R}^3$ evolve in the camera frame as $\dot{q} = \widehat{\omega} q + v$. Here, for $\omega \in \mathbb{R}^3$, $\widehat{\omega} \in so(3)$ is the skew-symmetric matrix generating the cross product. Then, after differentiating equation (4), we obtain:

$$\dot{\lambda}\mathbf{b} + \lambda\dot{\mathbf{b}} = \lambda\widehat{\omega}\mathbf{b} + v, \qquad (6)$$

where $\lambda = -e_3^T q + \xi r$, $e_3 \triangleq (0,0,1)^T$ and $r \triangleq \|q\|$. Now, using $q = \lambda\mathbf{b}$, we get $r = \lambda(1+e_3^T\mathbf{b})/\xi$. Also, it is clear that $\dot{\lambda} = -e_3^T(\widehat{\omega}q + v) + \xi q^T v/r$. Thus, after replacing all these expressions into (6), we obtain the following expression for the velocity of the back-projection ray in terms of the relative 3D camera motion:

$$\dot{\mathbf{b}} = -(I+\mathbf{b}e_3^T)\widehat{\mathbf{b}}\omega + \frac{1}{\lambda}\left(I + \mathbf{b}e_3^T - \frac{\xi^2\mathbf{b}\mathbf{b}^T}{1+e_3^T\mathbf{b}}\right)v. \qquad (7)$$

Since the first two components of the back-projection ray are simply $(x,y)^T$, the first two rows of (7) give us the expression for *central panoramic optical flow:*

$$\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} = \begin{bmatrix} xy & z-x^2 & -y \\ -(z-y^2) & -xy & x\end{bmatrix}\omega + \frac{1}{\lambda}\begin{bmatrix} 1-\rho x^2 & -\rho xy & (1-\rho z)x \\ -\rho xy & 1-\rho y^2 & (1-\rho z)y\end{bmatrix} v, \qquad (8)$$

where $\lambda = -Z + \xi\sqrt{X^2+Y^2+Z^2}$, $z = \mathbf{f}_\xi(x,y)$, and $\rho \triangleq \xi^2/(1+z)$. Notice that when $\xi = 0$, then $\rho = 0$ and (8) becomes the well-known equation for the optical flow of a perspective camera. When $\xi = 1$, then $\rho = 1/(x^2+y^2)$, and (8) becomes the equation for the optical flow of a paracatadioptric camera [10].

III. Multibody Motion Estimation and Segmentation

In this section, we derive a factorization method for estimating the motion of an unknown number n_o of independently moving objects from multiple central panoramic views. To this end, let $(x_i, y_i)^T$, $i = 1, \ldots, N$, be a pixel in the zeroth frame associated with one of the moving objects and let $(\dot{x}_{if}, \dot{y}_{if})^T$ be its optical flow in frame $f = 1, ..., F$, relative to the zeroth frame. We assume no prior segmentation of the image points, *i.e.* we do not know which image points correspond to which moving object. We define the *optical flow* matrix $W \in \mathbb{R}^{2N\times F}$ as:

$$W \triangleq \left[\begin{array}{ccc|ccc} \dot{x}_{11} & \cdots & \dot{x}_{N1} & \dot{y}_{11} & \cdots & \dot{y}_{N1} \\ \vdots & & \vdots & \vdots & & \vdots \\ \dot{x}_{1F} & \cdots & \dot{x}_{NF} & \dot{y}_{1F} & \cdots & \dot{y}_{NF}\end{array}\right]^T \qquad (9)$$

and the *segmentation matrix* $W_s \in \mathbb{R}^{N\times 2F}$ as:

$$W_s \triangleq \left[\begin{array}{cc|cc|cc} \dot{x}_{11} & \dot{y}_{11} & \cdots & \cdots & \dot{x}_{1F} & \dot{y}_{1F} \\ \vdots & \vdots & & & \vdots & \vdots \\ \dot{x}_{N1} & \dot{y}_{N1} & \cdots & \cdots & \dot{x}_{NF} & \dot{y}_{NF}\end{array}\right]. \qquad (10)$$

A. Single Body Motion Estimation

First consider the case where there is a single rigid body motion generating the optical flow, *i.e.* $n_o = 1$. Following equation (8), define the matrix of *rotational flows* Ψ and the matrix of *translational flows* Φ as:

$$\Psi = \begin{bmatrix} \{xy\} & \{z-x^2\} & -\{y\} \\ -\{z-y^2\} & -\{xy\} & \{x\}\end{bmatrix} \in \mathbb{R}^{2N\times 3},$$

$$\Phi = \begin{bmatrix} \{\frac{1-\rho x^2}{\lambda}\} & \{\frac{-\rho xy}{\lambda}\} & \{\frac{(1-\rho z)x}{\lambda}\} \\ \{\frac{-\rho xy}{\lambda}\} & \{\frac{1-\rho y^2}{\lambda}\} & \{\frac{(1-\rho z)y}{\lambda}\}\end{bmatrix} \in \mathbb{R}^{2N\times 3},$$

where, *e.g.* $\{xy\} = (x_1y_1, \cdots, x_Ny_N)^T \in \mathbb{R}^N$.

Then, the *optical flow* matrix $W \in \mathbb{R}^{2N\times F}$ satisfies [10]:

$$W = [\Psi \;\; \Phi]_{2N\times 6}\begin{bmatrix}\omega_1 & \cdots & \omega_F \\ v_1 & \cdots & v_F\end{bmatrix}_{6\times F} = SM^T, \qquad (11)$$

where ω_f and v_f are the rotational and linear velocities, respectively, of the object relative to the camera between the zeroth and the f-th frames. We call $S \in \mathbb{R}^{2N\times 6}$ the *structure* matrix and $M \in \mathbb{R}^{F\times 6}$ the *motion* matrix.

We conclude that the optical flow matrix generated by a single moving body undergoing a general translation and rotation has rank 6. Such a rank constraint can be naturally used to derive a factorization method for estimating the relative velocities (ω_f, v_f) and the scales λ_i from image points $(x_i, y_i)^T$ and optical flows $(\dot{x}_{if}, \dot{y}_{if})^T$. We can do so by factorizing W into its motion and structure components. To this end, consider the singular value decomposition (SVD) of $W = \mathcal{U}\mathcal{S}\mathcal{V}^T$, with $\mathcal{U} \in \mathbb{R}^{2N\times 6}$ and $\mathcal{V}\mathcal{S} \in \mathbb{R}^{F\times 6}$, 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: $\tilde{S}A_{1-3} = \Psi$ and $\tilde{S}A_{4-6} = \Phi$. Since Ψ is known, A_{1-3} can be immediately computed. The remaining columns of A and the vector of inverse scales $\{1/\lambda\} \in \mathbb{R}^N$ can be obtained up to scale from:

$$\begin{bmatrix} \mathrm{diag}(\{1-\rho x^2\}) & -\tilde{S}_x & 0 & 0 \\ -\mathrm{diag}(\{\rho xy\}) & 0 & -\tilde{S}_x & 0 \\ \mathrm{diag}(\{(1-\rho z)x\}) & 0 & 0 & -\tilde{S}_x \\ -\mathrm{diag}(\{\rho xy\}) & -\tilde{S}_y & 0 & 0 \\ \mathrm{diag}(\{1-\rho y^2\}) & 0 & -\tilde{S}_y & 0 \\ \mathrm{diag}(\{(1-\rho z)y\}) & 0 & 0 & -\tilde{S}_y \end{bmatrix} \begin{bmatrix} \{1/\lambda\} \\ A_4 \\ A_5 \\ A_6 \end{bmatrix} = 0.$$

where $\tilde{S}_x \in \mathbb{R}^{N\times 6}$ and $\tilde{S}_y \in \mathbb{R}^{N\times 6}$ are the upper and lower part of $\tilde{S}$, respectively.

B. Multibody Motion Segmentation

Let us now consider the case of a dynamic scene with n_o independent motions. In this case, the algorithm will proceed in two steps: (1) Use the *segmentation matrix* to associate groups of image measurements that correspond to the same moving object; (2) Given the segmentation, apply the algorithm in Section III-A to each group of measurements to estimate the motion of each object.

We first notice from (8) that, for a single motion k, the *segmentation matrix* $W_s \in \mathbb{R}^{N\times 2F}$ can be decomposed as

$$W_s = S_k M_k^T \tag{12}$$

where the structure matrix $S_k \in \mathbb{R}^{N\times 10}$ is given by

$$S_k = [\{xy\}, \{z-x^2\}, -\{y\}, -\{z-y^2\}, \{x\}, \{\frac{1-\rho x^2}{\lambda}\}, -\{\frac{\rho xy}{\lambda}\}, \{\frac{(1-\rho z)x}{\lambda}\}, \{\frac{1-\rho y^2}{\lambda}\}, \{\frac{(1-\rho z)y}{\lambda}\}]$$

and the motion matrix $M_k \in \mathbb{R}^{2F\times 10}$ is given by

$$M_k = \begin{bmatrix} \omega_{x1} & \omega_{y1} & \omega_{z1} & 0 & 0 & v_{x1} & v_{y1} & v_{z1} & 0 & 0 \\ -\omega_{y1} & 0 & 0 & \omega_{x1} & \omega_{z1} & 0 & v_{x1} & 0 & v_{y1} & v_{z1} \\ \vdots & & & & & & & & & \vdots \\ \omega_{xF} & \omega_{yF} & \omega_{zF} & 0 & 0 & v_{xF} & v_{yF} & v_{zF} & 0 & 0 \\ -\omega_{yF} & 0 & 0 & \omega_{xF} & \omega_{zF} & 0 & v_{xF} & 0 & v_{yF} & v_{zF} \end{bmatrix}.$$

Hence, for a single object in the scene, the collection of central panoramic optical flows across multiple frames lies on a ten-dimensional subspace of $\mathbb{R}^{2F}$.

More generally, the segmentation matrix associated with n_o independently moving objects can be decomposed as:

$$W_s = \begin{bmatrix} S_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S_{n_o} \end{bmatrix} \begin{bmatrix} M_1^T \\ \vdots \\ M_{n_o}^T \end{bmatrix} = SM^T$$

where $S \in \mathbb{R}^{N\times 10n_o}$, $M \in \mathbb{R}^{2F\times 10n_o}$, $S_k \in \mathbb{R}^{N_k\times 10}$, $M_k \in \mathbb{R}^{2F\times 10}$, N_k is the number of pixels associated with object k for $k = 1, \ldots, n_o$, and $N = \sum_{k=1}^{n_o} N_k$.

Since we are assuming that the segmentation of the image points is unknown, the rows of W_s may be in a different order. However, the reordering of the rows of W_s will not affect its rank. Assuming that $N \geq 10n_o$ and that $F \geq 5n_o$, we conclude that when the objects undergo a general motion the number of independent motions n_o can be estimated as $n_o = \mathrm{rank}(W_s)/10$.

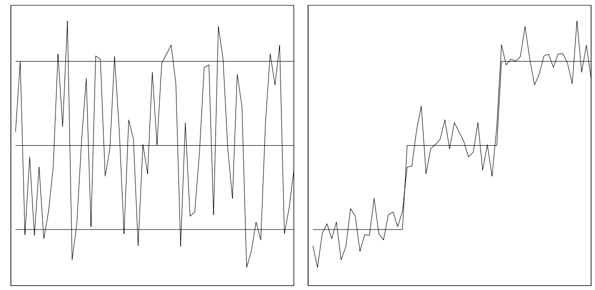

Fig. 3. Showing eigenvector based segmentation. Left: Elements of first eigenvector of a noisy W_s matrix. Right: Re-ordering the elements based on the levels.

One can show [14], [19] that the block diagonal structure of W_s implies that the entries of its leading singular vector will be the same for pixels corresponding to the same motion, and different otherwise. Thus, as proposed in [14], one can use the entries of the leading singular vector of W_s as an effective criterion for segmenting the pixels of the current frame into a collection of n_o groups corresponding to the independent motions. Figure 3 shows the elements of the leading singular vector for a noisy W_s matrix, and its re-ordering after thresholding its entries.

Once the optical flow has been segmented into independent motions, we can apply the factorization algorithm described in Section III-A separately on each group of measurements to estimate the motion of each object.

IV. Robot Navigation: Motion in X-Y Plane

The algorithm for infinitesimal motion estimation and segmentation described in Section III assumes that the motion of the objects is general, so that the subspace associated with each object is fully ten-dimensional. However, there are important robotics applications, *e.g.* ground robot navigation, in which the motion of the robots is restricted, hence the general algorithm described in Section III can not be directly applied.

In this section, we consider the special case where the moving objects are restricted to move only in the X-Y plane. In this case, the angular velocity is $\omega = (0, 0, \omega_z)^T$ and the linear velocity is $v = (v_x, v_y, 0)^T$. Hence, the optical flow generated by a single moving body is:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix} \omega_z + \frac{1}{\lambda} \begin{bmatrix} 1-\rho x^2 & -\rho xy \\ -\rho xy & 1-\rho y^2 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix}. \tag{13}$$

where $\lambda = -Z + \xi\sqrt{X^2+Y^2+Z^2}$, $Z = Z_{\mathrm{ground}}$, z is given in (5), and $\rho \triangleq \xi^2/(1+z)$.

A. Motion Estimation in the X-Y Plane

The optical flow matrix for a single body moving with velocities $\omega_f = (0, 0, \omega_{zf})^T$, $v_f = (v_{xf}, v_{yf}, 0)^T$ in frame f can be written as

$$W = [\Psi\ \Phi]_{2N\times 3} \begin{bmatrix} \omega_{z1} & \cdots & \omega_{zF} \\ v_{x1} & \cdots & v_{xF} \\ v_{y1} & \cdots & v_{yF} \end{bmatrix}_{3\times F} = SM^T,$$

with the matrices of rotational flows $\Psi \in \mathbb{R}^{2N\times 1}$ and translational flows $\Phi \in \mathbb{R}^{2N\times 2}$ given by

$$\Psi = \begin{bmatrix} -\{y\} \\ \{x\} \end{bmatrix}, \quad \Phi = \begin{bmatrix} \{\frac{1-\rho x^2}{\lambda}\} & \{\frac{-\rho xy}{\lambda}\} \\ \{\frac{-\rho xy}{\lambda}\} & \{\frac{1-\rho y^2}{\lambda}\} \end{bmatrix}.$$

We conclude that, in the case of a single object undergoing general motion in the X-Y plane, the optical flow matrix satisfies rank$(W) = 3$. Therefore, as in Section III-A, we consider the SVD of $W = \mathcal{U}\mathcal{S}\mathcal{V}^T$ and let $\tilde{S} = \mathcal{U} \in \mathbb{R}^{2N\times 3}$ and $\tilde{M} = \mathcal{V}\mathcal{S} \in \mathbb{R}^{F\times 3}$. Then $S = \tilde{S}A$ and $M = \tilde{M}A^{-T}$ for some $A \in \mathbb{R}^{3\times 3}$. Let A_k be the k-th column of A. Then the columns of A must satisfy: $\tilde{S}A_1 = \Psi$ and $\tilde{S}A_{2-3} = \Phi$. Since Ψ is known, A_1 can be immediately computed. The remaining columns of A and the vector of inverse scales $\{1/\lambda\} \in \mathbb{R}^n$ can be obtained up to scale from:

$$\begin{bmatrix} \text{diag}(\{1-\rho x^2\}) & -\tilde{S}_x & 0 \\ -\text{diag}(\{\rho xy\}) & 0 & -\tilde{S}_x \\ -\text{diag}(\{\rho xy\}) & -\tilde{S}_y & 0 \\ \text{diag}(\{1-\rho y^2\}) & 0 & -\tilde{S}_y \end{bmatrix} \begin{bmatrix} \{1/\lambda\} \\ A_2 \\ A_3 \end{bmatrix} = 0,$$

where $\tilde{S}_x \in \mathbb{R}^{N\times 3}$ and $\tilde{S}_y \in \mathbb{R}^{N\times 3}$ are the upper and lower part of $\tilde{S}$, respectively.

B. Motion Segmentation in the X-Y Plane

We now specialize the procedure of Section III-B to the case of planar motion. Following (13), the structure and motion matrices become

$$S_k = [-\{y\}, \{x\}, \{\frac{1-\rho x^2}{\lambda}\}, -\{\frac{\rho xy}{\lambda}\}, \{\frac{1-\rho y^2}{\lambda}\}] \in \mathbb{R}^{N\times 5}$$

$$M_k = \begin{bmatrix} \omega_{zf} & 0 & v_{xf} & v_{yf} & 0 \\ 0 & \omega_{zf} & 0 & v_{xf} & v_{yf} \\ \vdots & & & & \vdots \\ \omega_{zf} & 0 & v_{xf} & v_{yf} & 0 \\ 0 & \omega_{zf} & 0 & v_{xf} & v_{yf} \end{bmatrix} \in \mathbb{R}^{2F\times 5}.$$

We conclude that the collection of central panoramic optical flows for a single object moving in the X-Y plane lies in a 5-dimensional subspace of $\mathbb{R}^{2F}$. Therefore, the number of independent motions can be estimated as $n_o = \text{rank}(W_s)/5$. Then, the segmentation of the image measurements can be obtained as before from the leading singular vector of W_s. Once the optical flow has been segmented into independent motions, we can use the single-body factorization algorithm of Section IV-A to separately estimate the motion and structure of each independently moving object.

V. Experiments

Here we evaluate the performance of the proposed multibody motion estimation and segmentation algorithm in the case where two independently moving mobile robots are viewed by a static paracatadioptric camera ($\xi = 1$). The robots are equipped with GPS sensors with an accuracy of 2cm. We use the robots' GPS measurements to evaluate the performance of our motion estimation algorithm.

We grabbed 18 images of size 240×240 pixels at a framerate of 5Hz. The optical flow was computed directly in the image plane using Black's algorithm available at http://www.cs.brown.edu/people/black/ignc.html.

Fig. 1 shows a sample of the motion segmentation based on the optical flow. On the left, the optical flow generated by the two moving robots is shown, and on the right is the segmentation of the pixels corresponding to the independent motions. The two moving robots are segmented very well from the static background.

Fig. 4 and Fig. 5 show the output of our motion estimation algorithm for the two moving robots. These figures plot the estimated translational (v_x, v_y) and rotational velocity ω_z for the robots as a function of time in comparison with the values obtained by the on-board GPS sensors. Fig. 6 shows the root mean squared error for the motion estimates of the two robots. The vision estimates of linear velocity are within 0.15 m/s of the GPS estimates. The vision estimates of angular velocity are more noisy than the estimates of linear velocity, because the optical flow due to rotation is smaller than the one due to translation.

VI. Conclusions and Future Work

We have presented an algorithm for infinitesimal motion estimation and segmentation from multiple central panoramic views. Our algorithm is a factorization approach based on the fact that optical flows generated by a rigidly moving object across many frames lie in orthogonal ten-dimensional subspaces of a higher-dimensional space. We presented experimental results that show that our algorithm can effectively segment and estimate the motion of multiple moving objects from multiple catadioptric views.

Future work will consider relaxing the constraint on the motion of the objects being fully ten-dimensional. While the case of five-dimensional motion in the X-Y plane was easily handled with minor modifications of the general algorithm, the case of planar motion in an arbitrary plane is much more complex as demonstrated in [16] for the case of a single motion. Future work will also consider the extension of the current algorithm to the case of uncalibrated central panoramic cameras.

VII. Acknowledgments

We thank the support of ONR grant N00014-00-1-0621.

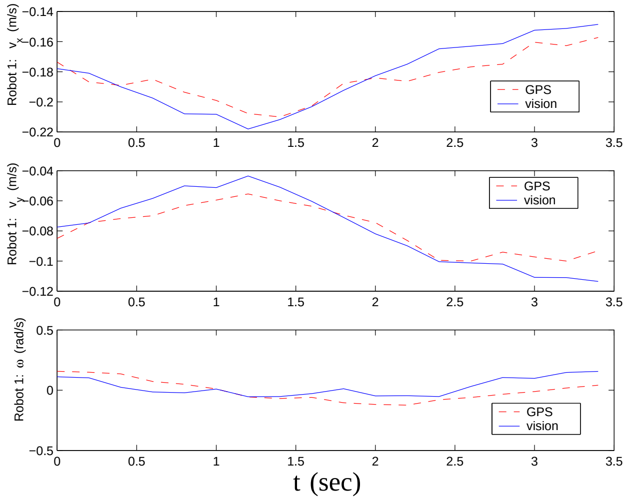

Fig. 4. Comparing the output of our vision-based motion estimation algorithm with GPS data for robot 1.

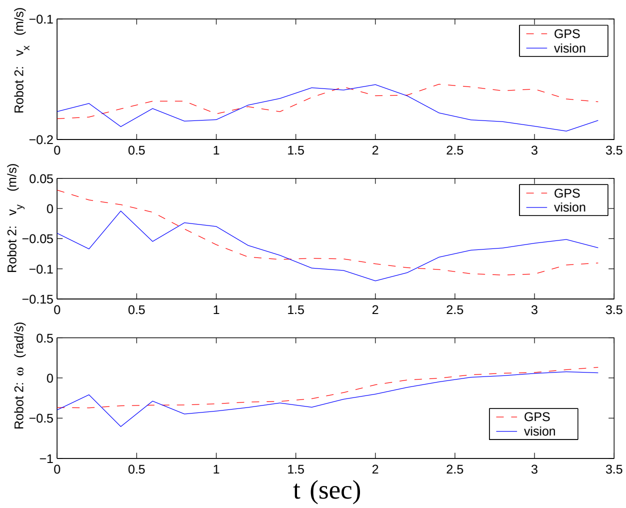

Fig. 5. Comparing the output of our vision-based motion estimation algorithm with GPS data for robot 2.

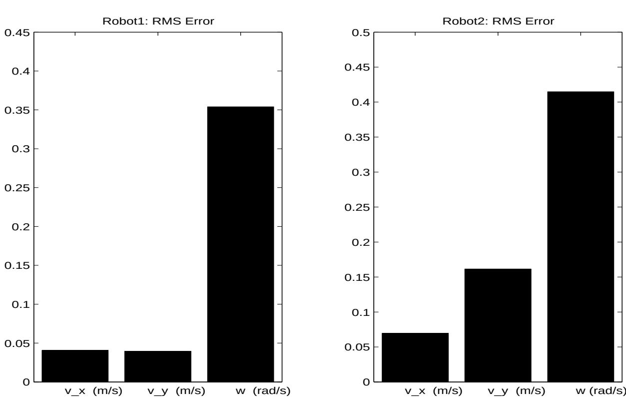

Fig. 6. Showing the RMS error for the motion estimates of the two robots.

VIII. REFERENCES

[1] S. Baker and S. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal on Computer Vision*, 35:175–196, 1999.

[2] J. Barreto and H. Araujo. Geometric properties of central catadioptric line images. In *7th European Conference on Computer Vision*, pages 237–251, May 2002.

[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] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In *European Conference on Computer Vision*, pages 445–461, 2000.

[5] C. Geyer and K. Daniilidis. Structure and motion from uncalibrated catadioptric views. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 279–286, 2001.

[6] C. Geyer and K. Daniilidis. Paracatadioptric camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(24):1–10, April 2002.

[7] J. Gluckman and S.K. Nayar. Ego-motion and omnidirectional cameras. In *Proceedings of IEEE 6th International Conference on Computer Vision*, pages 999–1005, 1998.

[8] M. Machline, L. Zelnik-Manor, and M. Irani. Multi-body segmentation: Revisiting motion consistency. In *Workshop on Vision and Modeling of Dynamic Scenes*, ECCV 2002.

[9] 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):206–18, 1997.

[10] O. Shakernia, R. Vidal, and S. Sastry. Infinitesimal motion estimation from multiple central panoramic views. In *IEEE Workshop on Motion and Video Computing*, pages 229–234, 2002.

[11] T. Svoboda, T. Pajdla, and V. Hlavac. Motion estimation using panoramic cameras. In *IEEE Conference on Intelligent Vehicles*, pages 335–350, 1998.

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

[13] R.F. Vassallo, J. Santos-Victor, and J. Schneebeli. A general approach for egomotion estimation with omnidirectional images. In *IEEE Workshop on Omnidirectional Vision*, pages 97–103, 2002.

[14] R. Vidal, A. Ames, and S. Sastry. Polynomial segmentation: an analytic solution to eigenvector segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2003. Submitted.

[15] R. Vidal, Y. Ma, S. Soatto, and S. Sastry. Two-view multibody structure from motion. *International Journal of Computer Vision*, 2002. Submitted.

[16] R. Vidal and J. Oliensis. Structure from planar motions with small baselines. In *European Conference on Computer Vision (LNCS 2351)*, volume 2, pages 383–398, 2002.

[17] R. Vidal, O. Shakernia, and S. Sastry. Formation control of nonholonomic mobile robots with omnidirectional visual servoing and motion segmentation. In *IEEE International Conference on Robotics and Automation*, 2003.

[18] R. Vidal, S. Soatto, and S. Sastry. A factorization method for multibody motion estimation and segmentation. In *Fortieth Annual Allerton Conference on Communication, Control and Computing*, pages 1625–1634, 2002.

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