Infinitesimal Motion Estimation from Multiple Central Panoramic Views*

Omid Shakernia   René Vidal   Shankar Sastry

Department of EECS, University of California, Berkeley, CA 94720
{omids,rvidal,sastry}@eecs.berkeley.edu

Abstract

We present an algorithm for infinitesimal motion estimation from multiple central panoramic views. We first derive the optical flow equations for central panoramic cameras as a function of both pixel coordinates and back-projection rays. We then derive a rank constraint on the optical flows across many frames, which must lie in a six dimensional subspace of a higher-dimensional space. We then propose factorization approach for recovering camera motion and scene structure. We present experimental results on a real image sequence.

Key words: Omni-directional vision, central panoramic optical flow, motion estimation, factorization methods.

1 Introduction

Omnidirectional vision systems provide a panoramic field of view, which can potentially greatly benefit the task of vision based motion estimation in applications such as mobile robot navigation. *Catadioptric cameras* [11] are realizations of omnidirectional vision systems which combine a mirror and a lens. *Central panoramic systems* are catadioptric cameras with a single effective viewpoint [1]. Examples of central panoramic systems with a unique effective viewpoint are a parabolic mirror in front of an orthographic camera, and a hyperbolic mirror in front of a perspective camera.

Motion estimation with omnidirectional vision has been an area of active research. Researchers have generalized many structure from motion algorithms from perspective projection to catadioptric projection in both the case of discrete motion [13, 6, 2] as well as differential motion [8, 16].

In [8, 16], 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 (*e.g.* Jepson and Heeger [9]) for spherical projection. In a more recent approach [3], the omnidirectional images are stereographically mapped onto the unit sphere and the image velocity field is computed on the sphere. Again, once the velocities are known on the sphere, one may apply any ego-motion algorithm for spherical projection.

In this paper, we derive the optical flow equations for central panoramic cameras *directly*, without the intermediate step of going to spherical projection. The optical flow equations that we derive depend not only on the pixel coordinates, but also on the back-projection rays. We then show that the optical flows accross many frames lie in a six dimensional subspace of a higher-dimensional space, generalizing the well-known *subspace constraints* [10] for perspective cameras to the case of central panoramic cameras. Further, we propose a factorization method to recover the infinitesimal camera motion and 3D structure based on multiple views. Our approach generalizes the work of [18], where subspace constraints were used to derive a factorization method for infinitesimal motion estimation in perspective cameras, to the case of central panoramic cameras.

Paper Outline: In Section 2 we describe the projection model for central panoramic cameras and derive the optical flow equations. In Section 3 we present an algorithm for estimating infinitesimal motion from multiple central panoramic views of a scene. In Section 4 we present experimental results, and we conclude in Section 5.

2 Central Panoramic Cameras

In this section, we describe the projection model for a central panoramic camera [5], and derive the central panoramic optical flow equation which relates the velocities of image points to infinitesimal camera motion.

The importance of the single viewpoint in central

*This research was supported by ONR grant N00014-00-1-0621.

panoramic cameras is that one can efficiently compute the *back-projection ray* (the ray from the optical center to the 3D point being imaged) associated with each image point. In fact, the optical flow equations we derive will be in terms of the image points and their associated back-projection rays.

We first present the special case of para-catadioptric cameras (a parabolic lens in front of an orthographic camera), then generalize to the case of any central panoramic camera.

2.1 Para-catadioptric Projection Model

A para-catadioptric camera first projects a 3D point onto the surface of a parabolic mirror and then orthographically projects onto an image plane (see Figure 1). The implicit equation for the surface of a parabola of focal length $1/2$ whose focus is at the origin is given by:

$$Z = \frac{1}{2}(X^2 + Y^2 - 1). \tag{1}$$

By intersecting a parameterized ray with the implicit equation of a paraboloid, and then orthographically projecting onto the image plane $Z = 0$, it can be shown that the image $\mathbf{x} = (x, y, 0)$ of a point $q = (X, Y, Z)^T$ is given by:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{-Z + \sqrt{X^2 + Y^2 + Z^2}} \begin{bmatrix} X \\ Y \end{bmatrix}. \tag{2}$$

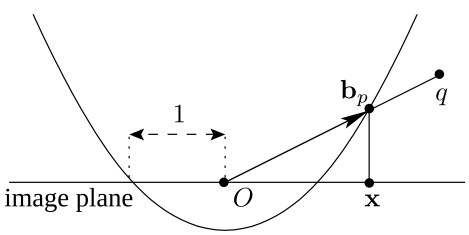

Figure 1. Showing the projection model for para-catadioptric cameras, and the back-projection ray $\mathbf{b}_p$ associated with image point $\mathbf{x}$.

The mapping from normalized image coordinates (x, y) to pixel coordinates (x_p, y_p) in the image involves an affine transformation:

$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} s_x & \sigma \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix},$$

where (s_x, s_y) are scales which combine information about the focal length of the mirror and the aspect ratio the orthographic projection, σ is the skew, and $(c_x, c_y)^T$ is the mirror center. In this paper, we assume that the catadioptric camera has been calibrated and work with normalized image coordinates. See, for example [7], for how to calibrate para-catadioptric cameras.

2.2 Para-catadioptric Optical Flow

We can write the para-catadioptric projection equation as:

$$\lambda \mathbf{x} = P q, \tag{3}$$

where $P = \text{diag}(1, 1, 0) \in \mathbb{R}^{3\times3}$, $\lambda = -e_3^T q + \|q\|$ is an unknown scale factor due to the para-catadioptric projection, and $e_3 = (0, 0, 1)^T$.

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 the static 3D point in the camera frame evolve as $\dot{q} = [\omega]_\times q + v$. Differentiating equation (3) with respect to time, we get $\dot{\lambda}\mathbf{x} + \lambda\dot{\mathbf{x}} = P([\omega]_\times q + v)$. Now, letting $r \triangleq \|q\|$, it is easy to show that

$$\dot{\lambda} = -e_3^T \dot{q} + \frac{q^T \dot{q}}{r} = -e_3^T([\omega]_\times q + v) + \frac{q^T v}{r}.$$

Then we have

$$\begin{aligned} \dot{\mathbf{x}} &= \frac{P(-[q]_\times \omega + v)}{\lambda} - \frac{\mathbf{x}}{\lambda}(-e_3^T(-[q]_\times \omega + v) + \frac{q^T v}{r}) \\ &= -(P + \mathbf{x} e_3^T)\frac{[q]_\times}{\lambda}\omega + \frac{1}{\lambda}(P + \mathbf{x} e_3^T - \frac{\mathbf{x} q^T}{r})v. \end{aligned}$$

Now, let $\mathbf{b}_p \in \mathbb{R}^3$ be the *back-projection ray* associated with image point $\mathbf{x}$ (see Figure 1). Using the implicit equation of the paraboloidal mirror (1), given an image point $\mathbf{x} = (x, y, 0)^T$ it is easy to compute its back-projection ray:

$$\mathbf{b}_p = (x, y, z)^T = (x, y, \frac{1}{2}(x^2 + y^2 - 1))^T. \tag{4}$$

In fact, since $P\mathbf{b}_p = \mathbf{x}$, and $\lambda\mathbf{x} = Pq$, we have $\mathbf{b}_p = q/\lambda$. Further, since $z = Z/\lambda$, and $\lambda = -Z + r$, we have $r = (1 + z)\lambda$. Hence, we obtain the following optical flow equation for para-catadioptric cameras:

$$\dot{\mathbf{x}} = -(P + \mathbf{x} e_3^T)[\mathbf{b}_p]_\times \omega + \frac{1}{\lambda}(P + \mathbf{x} e_3^T - \frac{\mathbf{x}\mathbf{b}_p^T}{1 + e_3^T \mathbf{b}_p})v. \tag{5}$$

Therefore, in terms of the back-projection ray $\mathbf{b}_p$ defined in equation (4), the optical flow $(\dot{x}, \dot{y})^T \triangleq (\mathrm{u}, \mathrm{v})^T$ induced by a para-catadioptric camera undergoing a motion (ω, v) is given by:

$$\begin{bmatrix} \mathrm{u} \\ \mathrm{v} \end{bmatrix} = \begin{bmatrix} xy & z - x^2 & -y \\ -(z - y^2) & -xy & x \end{bmatrix} \omega + \frac{1}{\lambda} \begin{bmatrix} \frac{1+z-x^2}{1+z} & \frac{-xy}{1+z} & \frac{x}{1+z} \\ \frac{-xy}{1+z} & \frac{1+z-y^2}{1+z} & \frac{y}{1+z} \end{bmatrix} v. \tag{6}$$

2.3 Central Panoramic Projection Model

It was shown in [5] that all central panoramic systems 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 (see Figure 2). The projection model defined in [5] represents in a unified manner all central panoramic systems with two parameters (l, m). For example, a para-catadioptric camera would have parameters $(l, m) = (1, 0)$, while a pinhole camera would have parameters $(l, m) = (0, 1)$.

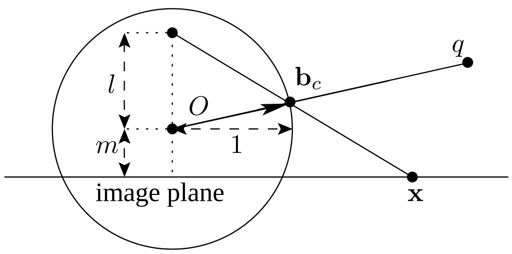

Figure 2. Showing the projection model for central panoramic cameras, and the back-projection ray $\mathbf{b}_c$ associated with image point $\mathbf{x}$.

The image $\mathbf{x} = (x, y, -m)^T$ of a 3D point $q = (X, Y, Z)^T$ obtained through a central panoramic system with parameters (l, m) is given by [5]:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{l+m}{-Z + l\sqrt{X^2+Y^2+Z^2}} \begin{bmatrix} X \\ Y \end{bmatrix}. \qquad (7)$$

As in the previous section, we work in normalized image coordinates, and assume that the camera has been previously calibrated.

2.4 Central Panoramic Optical Flow

We can write the central panoramic projection equation as:

$$\lambda(\mathbf{x} + m e_3) = P q \qquad (8)$$

where $P = \text{diag}(1, 1, 0)$, $e_3 = (0, 0, 1)^T$, and $\lambda = \frac{1}{l+m}(-e_3^T q + l\|q\|)$ is an unknown scale factor due to the central panoramic projection.

If the camera undergoes a linear velocity $v \in \mathbb{R}^3$ and an angular velocity $\omega \in \mathbb{R}^3$, then we have $\dot{q} = [\omega]_\times q + v$. Differentiating (8) with respect to time, we get

$$\dot{\lambda}(\mathbf{x} + m e_3) + \lambda \dot{\mathbf{x}} = P([\omega]_\times q + v).$$

Noticing that $\mathbf{x} + m e_3 = P\mathbf{x}$, we have

$$\dot{\mathbf{x}} = \frac{P([\omega]_\times q + v) - \dot{\lambda} P \mathbf{x}}{\lambda}.$$

Now, letting $r \triangleq \|q\|$, and $\mu \triangleq (l + m)$, it is easy to show that

$$\dot{\lambda} = \frac{1}{\mu}\left(-e_3^T([\omega]_\times q + v) + l\frac{q^T v}{r}\right)$$

and

$$\dot{\mathbf{x}} = -\frac{1}{\lambda} P(I + \frac{\mathbf{x} e_3^T}{\mu})[q]_\times \omega + \frac{1}{\lambda} P(I + \frac{\mathbf{x}}{\mu}(e_3^T - \frac{l q^T}{r}))v. \quad (9)$$

Now, let $\mathbf{b}_c = (b_1, b_2, b_3)^T \in \mathbb{R}^3$ be the *back-projection ray* associated with image point $\mathbf{x} = (x, y, -m)^T$ (see Figure 2). By intersecting the unit sphere with the line segment connecting points $\mathbf{x}$ and $(0, 0, l)^T$, we can compute $\mathbf{b}_c$ as:

$$\begin{aligned} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &= \frac{l\mu + \mathtt{sign}(\mu)\sqrt{(x^2+y^2)(1-l^2)+\mu^2}}{x^2+y^2+\mu^2} \begin{bmatrix} x \\ y \end{bmatrix} \\ b_3 &= \pm\sqrt{1 - b_1^2 - b_2^2} \end{aligned}$$

where b_3 is negative if $|\mu/l| > \sqrt{x^2+y^2}$ and positive otherwise. Now, since $\|\mathbf{b}_c\| = 1$, we have $\mathbf{b}_c = q/r$, which we substitute into equation (9) to get

$$\dot{\mathbf{x}} = -\frac{r}{\lambda} P(I + \frac{\mathbf{x} e_3^T}{\mu})[\mathbf{b}_c]_\times \omega + P(\frac{I}{\lambda} + \frac{\mathbf{x}(e_3^T - l\mathbf{b}_c^T)}{\lambda\mu})v.$$

Now, using the relations $b_3 = Z/r$ and $\lambda = (-Z + lr)/\mu$ we have $r/\lambda = \mu/(l - b_3)$. Therefore we have derived the following optical flow equation for central panoramic systems:

$$\dot{\mathbf{x}} = -P(\frac{\mu I + \mathbf{x} e_3^T}{l - e_3^T \mathbf{b}_c})[\mathbf{b}_c]_\times \omega + P(\frac{I}{\lambda} + \frac{\mathbf{x}(e_3^T - l\mathbf{b}_c^T)}{\lambda\mu})v. \quad (10)$$

Therefore in terms of the back-projection ray $\mathbf{b}_c$, the optical flow $(\dot{x}, \dot{y})^T \triangleq (\mathrm{u}, \mathrm{v})^T$ induced by a central panoramic camera undergoing a motion (ω, v) is given by:

$$\begin{aligned} \begin{bmatrix} \mathrm{u} \\ \mathrm{v} \end{bmatrix} = &\begin{bmatrix} \frac{x b_2}{l - b_3} & \frac{-x b_1 + \mu b_3}{l - b_3} & \frac{-\mu b_2}{l - b_3} \\ \frac{y b_2 - \mu b_3}{l - b_3} & \frac{-y b1}{l - b_3} & \frac{\mu b_1}{l - b_3} \end{bmatrix} \omega + \qquad (11) \\ &\frac{1}{\lambda} \begin{bmatrix} \frac{\mu - x l b_1}{\mu} & \frac{-x l b_2}{\mu} & \frac{x(1 - l b_3)}{\mu} \\ \frac{-y l b_1}{\mu} & \frac{\mu - y l b_2}{\mu} & \frac{y(1 - l b_3)}{\mu} \end{bmatrix} v. \end{aligned}$$

Comments 2.1. *Since para-catadioptric projection is a special case of the central panoramic projection, one would expect that plugging values $(l, m) = (1, 0)$ into central panoramic optical flow equation (10) would yield the para-catadioptric optical flow equation (5). Although by the form of the equations, this appears not to be the case, in fact it is. Notice that in the para-catadioptric case, we had $\mathbf{b}_p = q/\lambda$, while in the central panoramic case, we have $\mathbf{b}_c = q/r$, and hence $\mathbf{b}_p = \frac{r}{\lambda}\mathbf{b}_c$. Also, in the para-catadioptric case $e_3^T \mathbf{x} = 0$ while in the central panoramic case $e_3^T \mathbf{x} = -m$. Using these facts, it is direct to check that the optical flow for the central panoramic case reduces to that of the para-catadioptric case when $(l, m) = (1, 0)$.*

3 Infinitesimal motion estimation from multiple central panoramic views

Tomasi and Kanade [14] proposed an algorithm to estimate the motion of an orthographic camera, based on discrete image measurements. They used a factorization method based on the fact that, under orthographic projection, discrete image measurements lie on a low-dimensional linear variety. The method has also been extended to affine and paraperspective cameras [12]. Unfortunately, under full perspective projection such a variety is nonlinear [15], hence factorization methods cannot be used. However, Irani [10] showed that the infinitesimal measurements do lie on a low-dimensional linear variety in the perspective projection case. Irani used subspace constraints on the motion field to obtain a multi-frame algorithm for the estimation of the optical flow of a moving camera observing a static scene. She did not use those constraints for 3D motion estimation.

Here, we generalize the results of [18], where subspace constraints were used to derive a factorization method for infinitesimal motion estimation in perspective cameras, to the case of central panoramic cameras.

Given measurements for the optical flow $(\dot{x}_j^i, \dot{y}_j^i)^T = (\mathrm{u}_j^i, \mathrm{v}_j^i)^T$ of point $i = 1, \ldots, n$ in frame $j = 1, \ldots, m$ relative to a reference frame, let

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

and define the *optical flow* matrix $W \in \mathbb{R}^{2n\times m}$:

$$W = \begin{bmatrix} U \\ V \end{bmatrix}.$$

3.1 Para-catadioptric Case

In the para-catadioptric case, let $\mathbf{b}_p^i = (x^i, y^i, z^i)^T$, for $i = 1, \ldots, n$, be the back-projection rays in the reference frame. Following equation (6), define the matrix of *rotational flows* Ψ_p and the matrix of *translational flows* Φ_p as:

$$\begin{aligned} \Psi_p &= \begin{bmatrix} \{xy\} & \{z - x^2\} & -\{y\} \\ -\{z - y^2\} & -\{xy\} & \{x\} \end{bmatrix} \in \mathbb{R}^{2n\times 3}, \\ \Phi_p &= \begin{bmatrix} \{\frac{1+z-x^2}{\lambda(1+z)}\} & \{\frac{-xy}{\lambda(1+z)}\} & \{\frac{x}{\lambda(1+z)}\} \\ \{\frac{-xy}{\lambda(1+z)}\} & \{\frac{1+z-y^2}{\lambda(1+z)}\} & \{\frac{y}{\lambda(1+z)}\} \end{bmatrix} \in \mathbb{R}^{2n\times 3}, \end{aligned}$$

where (for example) $\{xy\} = (x^1y^1, \cdots, x^ny^n)^T \in \mathbb{R}^n$.

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

$$W = [\Psi_p \;\; \Phi_p]_{2n\times 6} \begin{bmatrix} \omega_1 & \cdots & \omega_m \\ v_1 & \cdots & v_m \end{bmatrix}_{6\times m} = SM^T, \quad (12)$$

where ω_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.

The rank constraint $\mathrm{rank}(W) = 6$ can be naturally used to derive a factorization method for estimating the relative velocities (ω_j, v_j) and scales λ^i from back-projection rays $\mathbf{b}_p^i$ and optical flows $\dot{\mathbf{x}}_j^i$. We can do so by factorizing W into its motion and structure components. For, consider the 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: $\tilde{S}A_{1-3} = \Psi_p$ and $\tilde{S}A_{4-6} = \Phi_p$. Since Ψ_p 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:

$$\begin{bmatrix} -\mathrm{diag}(\{(\frac{1+z-x^2}{1+z}\}) & \tilde{S}_{\mathrm{u}} & 0 & 0 \\ \mathrm{diag}(\{\frac{xy}{1+z}\}) & 0 & \tilde{S}_{\mathrm{u}} & 0 \\ -\mathrm{diag}(\{\frac{x}{1+z}\}) & 0 & 0 & \tilde{S}_{\mathrm{u}} \\ \mathrm{diag}(\{\frac{xy}{1+z}\}) & \tilde{S}_{\mathrm{v}} & 0 & 0 \\ -\mathrm{diag}(\{(\frac{1+z-y^2}{1+z}\}) & 0 & \tilde{S}_{\mathrm{v}} & 0 \\ -\mathrm{diag}(\{\frac{y}{1+z}\}) & 0 & 0 & \tilde{S}_{\mathrm{v}} \end{bmatrix} \begin{bmatrix} \{1/\lambda\} \\ A_4 \\ A_5 \\ A_6 \end{bmatrix} = 0.$$

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.

3.2 Central Panoramic Case

In the case of general central panoramic cameras, following equation (11), we define Ψ_c and Φ_c as:

$$\begin{aligned} \Psi_c &= \begin{bmatrix} \{\frac{xb_2}{l-b_3}\} & \{\frac{-xb_1+\mu b_3}{l-b3}\} & \{\frac{-\mu b_2}{l-b3}\} \\ \{\frac{yb_2-\mu b_3}{l-b3}\} & \{\frac{-yb1}{l-b3}\} & \{\frac{\mu b_1}{l-b3}\} \end{bmatrix} \in \mathbb{R}^{2n\times 3}, \\ \Phi_c &= \begin{bmatrix} \{\frac{\mu - xlb_1}{\lambda\mu}\} & \{\frac{-xlb_2}{\lambda\mu}\} & \{\frac{x(1-lb_3)}{\lambda\mu}\} \\ \{\frac{-ylb_1}{\lambda\mu}\} & \{\frac{\mu - ylb_2}{\lambda\mu}\} & \{\frac{y(1-lb_3)}{\lambda\mu}\} \end{bmatrix} \in \mathbb{R}^{2n\times 3}. \end{aligned}$$

As described in Section 3.1, we can factorize the optical flow matrix W into into its motion and structure components. For, consider the 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: $\tilde{S}A_{1-3} = \Psi_c$ and $\tilde{S}A_{4-6} = \Phi_c$. Since Ψ_c 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:

$$\begin{bmatrix} \mathrm{diag}(\{\frac{xlb_1-\mu}{\mu}\}) & \tilde{S}_{\mathrm{u}} & 0 & 0 \\ \mathrm{diag}(\{\frac{xlb_2}{\mu}\}) & 0 & \tilde{S}_{\mathrm{u}} & 0 \\ \mathrm{diag}(\{\frac{x(lb_3-1)}{\mu}\}) & 0 & 0 & \tilde{S}_{\mathrm{u}} \\ \mathrm{diag}(\{\frac{ylb_1}{\mu}\}) & \tilde{S}_{\mathrm{v}} & 0 & 0 \\ \mathrm{diag}(\{\frac{ylb_2-\mu}{\mu}\}) & 0 & \tilde{S}_{\mathrm{v}} & 0 \\ \mathrm{diag}(\{\frac{y(lb_3-1)}{\mu}\}) & 0 & 0 & \tilde{S}_{\mathrm{v}} \end{bmatrix} \begin{bmatrix} \{1/\lambda\} \\ A_4 \\ A_5 \\ A_6 \end{bmatrix} = 0.$$

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.

3.3 Robot Navigation: Motion in the X-Y plane

Consider the special case where the camera is restricted to move only in the X-Y plane. This is the case shows up, for example, when a camera is attached to a mobile robot (with the optical axis parallel to the Z axis) and the robot can only rotate and translate along the ground plane. In this case, the angular velocity is $\omega = (0, 0, \omega_3)^T$, the linear velocity is $v = (v_1, v_2, 0)^T$. Hence in the para-catadioptric case (a similar simplification arises in the general central panoramic case), we have:

$$\begin{bmatrix} \mathrm{u} \\ \mathrm{v} \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix} \omega_3 + \frac{1}{\lambda} \begin{bmatrix} \frac{1+z-x^2}{1+z} & \frac{-xy}{1+z} \\ \frac{-xy}{1+z} & \frac{1+z-y^2}{1+z} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}. \quad (13)$$

With this simplification, the we can apply an analogous algorithm to that described in Section 3.1, except that here the matrices of rotational flows $\Psi \in \mathbb{R}^{2n\times 1}$ and translational flows $\Phi \in \mathbb{R}^{2n\times 2}$ are given by

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

with

$$W = [\Psi\ \Phi]_{2n\times 3} \begin{bmatrix} \omega_{3_1} & \cdots & \omega_{3_m} \\ v_{1_1} & \cdots & v_{1_m} \\ v_{2_1} & \cdots & v_{2_m} \end{bmatrix}_{3\times m} = SM^T$$

where motion (ω_j, v_j) in frame j is $\omega_j = (0, 0, \omega_{3_j})^T$ and $v_j = (v_{1_j}, v_{2_j}, 0)^T$. We conclude that the optical flow satisfies $\mathrm{rank}(W) = 3$.

As described in Section 3.1, we can factorize the optical flow matrix W into into its motion and structure components. For, consider the 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 $A \in \mathbb{R}^{3\times 3}$. Let A_k be the k-th column of A. Then the columns of A must satisfy: $\tilde{S}A_1 = \Psi_c$ 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:

$$\begin{bmatrix} -\mathrm{diag}(\{(\frac{1+z-x^2}{1+z}\}) & \tilde{S}_{\mathrm{u}} & 0 \\ \mathrm{diag}(\{\frac{xy}{1+z}\}) & 0 & \tilde{S}_{\mathrm{u}} \\ \mathrm{diag}(\{\frac{xy}{1+z}\}) & \tilde{S}_{\mathrm{v}} & 0 \\ -\mathrm{diag}(\{(\frac{1+z-y^2}{1+z}\}) & 0 & \tilde{S}_{\mathrm{v}} \end{bmatrix} \begin{bmatrix} \{1/\lambda\} \\ A_2 \\ A_3 \end{bmatrix} = 0.$$

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.

4 Experimental Results

Here we evaluate the performance of the proposed motion estimation algorithm in the case where a nonholonomic mobile robot moving in the X-Y plane is viewed by a static para-catadioptric camera. The robot is equipped with GPS sensors with an accuracy of 2cm. We use the GPS measurements as the ground truth with which we evaluate the performance of our motion estimation algorithm. One of the seven frames of the image sequence is shown in Figure 3.

The optical flow (shown on the left column of Figure 3), was computed using Black's algorithm available at http://www.cs.brown.edu/people/black/ignc.html. We first used the optical flow to segment the pixels corresponding to the moving object from those of the background. This was done simply by looking at pixels for which the norm of the optical flow is larger than a threshold. The right column of Figure 3 shows the motion segmentation result. Only the pixels in the images with large enough flow vectors were used to estimate the motion of the robot.

Figure 4 plots the ground truth as well as the estimated rotational velocity ω_z and translational velocity (v_x, v_y) for each frame. Figure 5 shows the root mean squared error for the motion estimates. Notice that the estimates for angular velocity are considerable more noisy than linear velocity. This is because not as much optical flow is generated when the robot rotates in the scene as compared to when it translates.

5 Conclusions and Future Research

We have presented an algorithm for infinitesimal motion estimation from multiple central panoramic views. Our algorithm is a factorization approach based on the fact that optical flows across many frames lie on a 6 dimensional subspace of a higher-dimensional space. We presented experimental results that show that our algorithm can effectively recover camera motion from multiple catadioptric views.

Future work will include extending our algorithm to motion segmentation and estimation for multiple independently moving objects. We also plan to apply our results to the problems of pursuit-evasion games [17], mobile robot navigation and formation control [4].

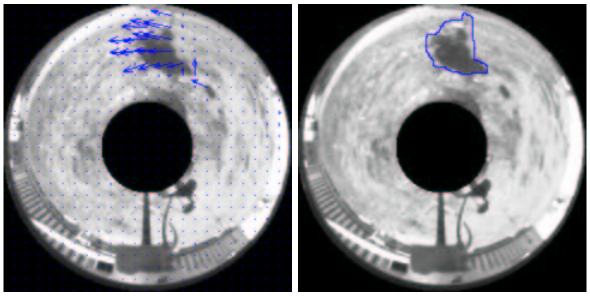

Figure 3. The optical flow and motion segmentation of a robot in an image sequence.

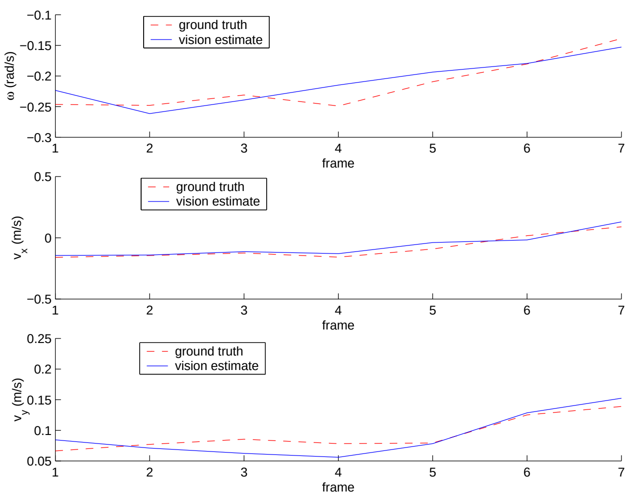

Figure 4. Motion estimation results.

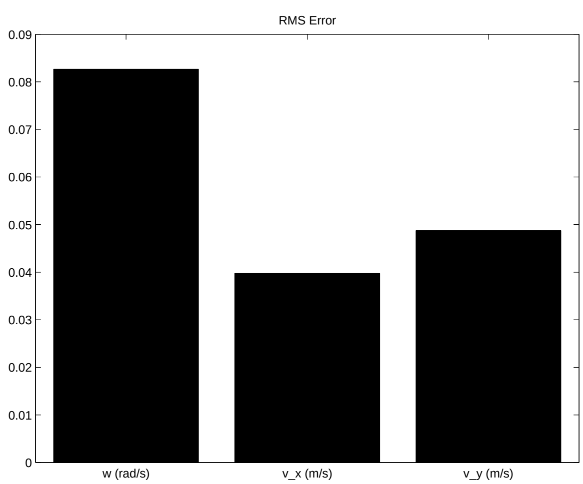

Figure 5. Root Mean Square Error of the motion estimates.

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] P. Chang and M. Hebert. Omni-directional structure from motion. In *IEEE Workshop on Omnidirectional Vision*, pages 127–133, June 2000.

[3] K. Daniilidis, A. Makadia, and T. Bulow. Image processing in catadioptric planes: Spatiotemporal derivatives and optical flow computation. In *Proceedings of IEEE Workshop on Omni-directional Vision*, pages 3–10, 2002.

[4] O. directional Vision-Based Distributed Formation Control of Nonholonomic Mobile Robots. René vidal and omid shakernia and shankar sastry. In *IEEE International Conference on Robotics and Automation*, 2003. Submitted.

[5] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In *Proc. of the European Conference on Computer Vision*, 2000.

[6] 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.

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

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

[9] D. Heeger and A. Jepson. Subspace methods for recovering rigid motion. *Int. J. on Computer Vision*, 7(2):95–117, 1992.

[10] M. Irani. Multi-frame optical flow estimation using subspace constraints. In *IEEE International Conference on Computer Vision*, volume 2, pages 626–633, 1999.

[11] S. Nayar. Catadioptric omnidirectional camera. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 482–488, Puerto Rico, 1997.

[12] 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.

[13] T. Svoboda, T. Pajdla, and V. Hlavac. Epipolar geometry for panoramic cameras. In *5th European Conference on Computer Vision*, pages 218–231, 1998.

[14] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography. *Int. J. on Computer Vision*, 9(2):137–154, 1992.

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

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

[17] R. Vidal, O. Shakernia, J. Kim, D. Shim, and S. Sastry. Probabilistic pursuit-evasion games: theory, implementation and experimental evaluation. *IEEE Transactions on Robotics and Automation*, 2002. To appear.

[18] R. Vidal, S. Soatto, and S. Sastry. A factorization method for 3D multi-body motion estimation and segmentation. In *Proceedings of 40th Annual Allerton Conference on Communication, Control, and Computing*, 2002. To appear.