Navigation using a spherical camera

Raman Arora
University Of Wisconsin-Madison

Harish Parthasarathy
Netaji Subhas Institute of Technology

Abstract

A novel group theoretical method is proposed for autonomous navigation based on a spherical image camera. The environment of a robot is captured on a sphere. The three dimensional scenes at two different points in the space are related by a transformation from the special Euclidean motion group which is the semi-direct product of the rotation and the translation groups. The motion of the robot is recovered by iteratively estimating the rotation and the translation in an Expectation-Maximization fashion.

1 Introduction

An iterative approach based on matched-filtering is presented for visual localization and navigation of a robot with an omnidirectional sensor. The focus is on the omni-directional imaging techniques that capture the three dimensional scene on the sphere. For instance, the catadioptric systems, which incorporate convex parabolic or hyperbolic reflectors with the camera, allow the captured images to be be mapped onto a regular spherical grid [7, 9]. The three dimensional scene may also be captured on the sphere using a pair of fisheye lenses [3]. Such vision systems find applications in robotics [6], video surveillance, medical imaging and automatic face recognition. Various localization and motion-estimation or motion-recovery techniques in robotics employ omni-directional imagery [4, 5, 6, 8].

In this paper we develop a method to recover transformations between two images defined on a sphere. Such a problem may arise when a robot or a flying drone needs to orient itself in space and navigate to a pre-determined goal position. The robot captures the omnidirectional scene on a sphere and compares it with a pre-stored spherical image to recover the possible rotation and translation connecting the two scenes. The robot may then follow the computed trajectory to the goal updating it as the scene along the path changes. Whereas a differential visual-homing technique was put forth in [8], we present a group-theoretic approach.

The problem of autonomous vision-based navigation is formulated in Section 2. The problem is presented as one of inferring image transformations from a pair of images defined on a manifold. The transformation comes from the special Euclidean motion group $SE(3)$ which is not compact but can be expressed as the semi-direct product of the rotation group $SO(3)$ (which is compact) and the translation group $\mathbb{R}^3$ (which is abelian). A matched filter is developed in Section 3 to recover the translation and Section 4 discusses the recovery of rotation component. Experimental results are presented in Section 5.

2 Problem formulation

Imagine a robot equipped with a spherical vision system exploring its environment. The omni-directional camera mounted on the robot captures the three dimensional scene at regular intervals along robot's path. The motion of robot induces a transformation of the observed omnidirectional scene, with transformation coming from the three dimensional special Euclidean motion group, $SE(3)$. The $SE(3)$ group is the semiproduct of $\mathbb{R}^3$ with the special orthogonal group $SO(3)$. An element $t \in SE(3)$ can be denoted as $t = (\mathbf{r}, R)$ where $\mathbf{r} \in \mathbb{R}^3$ and $R \in SO(3)$. The objective is to determine Euclidean motion connecting any two points in space, given the spherical images at those locations. An expectation-maximization approach is presented in this paper wherein the rotation is estimated in the expectation step and in the maximization step the de-rotated images are matched-filter to recover the relative depth or translation.

A point on the unit sphere is given, in polar coordinates, by the pair of angles (θ, ϕ) where $\theta \in [0, \pi]$ represents the colatitude and $\phi \in [0, 2\pi)$ is the azimuth. An infinitesimal small region of the scene at depth z with locally-planar brightness distribution $f(x, y)$ can be regarded as a patch on the unit sphere with brightness distribution in angular coordinates given by $\xi(\theta, \phi) = f(\ T(z, \theta, \phi)\ ) + v(\theta, \phi)$, for $\theta \in [0, \pi]$ and

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

$\phi \in [0, 2\pi)$, where T is a fundamental map that determines the projection of the scene on the sphere and is determined by the specifics of the imaging device. Note that a Euclidean motion acts on ξ through T.

The captured image ξ is considered to be a random field defined on the unit sphere, comprising of a signal component and a noise component $\xi(x) = s(x) + v(x),\ x \in S^2$. The noise field is assumed to be isotropic, i.e. the correlation function $K(x_1, x_2) = \mathbb{E}\left[v(x_1)v(x_2)\right]$ satisfies the rotational invariance: $K(g \cdot x_1, g \cdot x_2) = K(x_1, x_2)$ for all $x_1, x_2 \in S^2$ and $g \in SO(3)$.

To recover translation, filters matched to the signal $\xi(\theta, \phi)$ are designed for various $z \in \{z_1, \dots, z_m\}$ and the recorded image is filtered with each of these filters. The z_i corresponding to the maximum SNR is the best estimate of the depth. When the spherical camera undergoes pure rotation, the two images $\xi^{(1)}, \xi^{(2)}$ are related as $\xi^{(2)}(x) = \xi^{(1)}(g^{-1} \cdot x)$ for $x \in S^2, g \in SO(3)$. The problem reduces to determining the rotation angle and the axis of rotation, from the spherical images $\xi^{(1)}$ and $\xi^{(2)}$. Given the true rotation of camera, the images $\xi^{(1)}, \xi^{(2)}$ after de-rotation are simply related by a pure translation. This observation motivates the EM algorithm for robust navigation. The translation is treated as an unobserved latent variable. Therefore, the expectation (E) step computes an estimate of the rotation angle by compensating for translation as if it were observed (initialize with no translation). The maximization (M) step computes the maximum likelihood estimates of the translation based on the spherical matched filter and the estimates of rotation from the E step. The translation parameters found on the M step are then used to begin another E step, and the process is repeated. The knowledge of fundamental map T is crucial to generate matched filters to estimate depth.

The problem is addressed using spherical harmonic transform: a square integrable function on the sphere can be expanded in the series [2],

$$\xi(x) = \sum_{n=0}^{\infty} \sum_{j=-n}^{n} d_n\ \hat{\xi}_{n,j}\ \psi_{n,j} \qquad (1)$$

where $\psi_{n,j}$ are spherical harmonics, $d_n = (2n+1)$, $\hat{\xi}_{n,j} = \langle \xi, \psi_{n,j} \rangle$ are spherical harmonic coefficients.

3 Depth from Spherical Matched filter

Let h denote a filter on the unit sphere S^2. Given the random field $\{\xi(x) : x \in S^2\}$ as input, the filtered output sampled at a fixed point $y \in S^2$ is given as

$$\eta(y) = \int_{S^2} h(x, y)\xi(x)\, dx \qquad (2)$$

The variable y will be suppressed henceforth for notational convenience. The signal and noise components of the filtered output are denoted as $\eta_s = \int_{S^2} h(x)s(x)dx$ and $\eta_v = \int_{S^2} h(x)v(x)dx$ respectively. The aim is to design the filter h that maximizes the signal-to-noise ratio $\Phi(h) = \frac{|\eta_s|^2}{\mathbb{E}|\eta_v|^2}$.

Let x_0 denote the north pole of the sphere and g be the rotation that takes north pole to x. Making substitutions $\theta(g) = s(gx_0)$ and $\phi(g) = h(gx_0)$ the signal component of the filtered output can be expressed as integral over $SO(3)$,

$$\eta_s = \int_{S^2} h(x)s(x)dx = \int_{SO(3)} \phi(g)\theta(g)dg. \qquad (3)$$

Similarly, the noise power is expressed as an integral on $SO(3) \times SO(3)$,

$$\begin{aligned}\mathbb{E}[|\eta_v|^2] &= \int_{S^2 \times S^2} K(x_1, x_2)h(x_1)h(x_2)dx_1dx_2 \\ &= \int_{SO(3) \times SO(3)} k(g_2^{-1}g_1)\phi(g_1)\phi(g_2)dg_1dg_2 \qquad (4)\end{aligned}$$

where $k(g) = K(gx_0, x_0)$. Using the Parseval's relation, the signal component and noise power can be expressed in spherical harmonic series (1),

$$\begin{aligned}\eta_s &= \sum_{n=0}^{\infty} \sum_{j=-n}^{n} d_n \hat{\phi}_{n,j} \hat{\theta}^*_{n,j} \\ \mathbb{E}[|\eta_v|^2] &= \sum_{n=0}^{\infty} \sum_{j=-n}^{n} d_n \lambda_n |\hat{\phi}_{n,j}|^2 \qquad (5)\end{aligned}$$

where $\lambda_n = \sum_{j=-n}^{n} \mathbb{E}\left[\langle v, \psi_{n,j} \rangle^2\right]$. λ_n can be interpreted as sum of the variances of noise component along each eigenvector of n^{th} irreducible subspace of representation of $SO(3)$. The matched filter is given by the function h that maximizes the signal to noise ratio,

$$\Phi(h) = \frac{|\sum_{n,j} d_n \hat{\phi}_{n,j} \hat{\theta}^*_{n,j}|^2}{\sum_{n,j} d_n \lambda_n |\hat{\phi}_{n,j}|^2}. \qquad (6)$$

Assuming λ_n's are non zero, the application of Cauchy-Schwarz inequality gives the upper bound on the SNR,

$$\Phi(h) \leq \sum_{n,j} d_n \lambda_n^{-1} |\hat{\theta}_{n,j}|^2 \qquad (7)$$

where equality holds if and only if $\hat{\phi}_{n,j} = \lambda_n^{-1}\hat{\theta}_{n,j}$.

Therefore, the matched filter is implemented by first calculating spherical harmonic coefficients of the template,

$$\hat{\theta}_{n,j} = \int_{S^2} s(x)\psi^*_{n,j}(x)dx \qquad (8)$$

and then computing

$$h(x) \; = \; \phi(\tau^{-1}(x)) = \sum_{n=0}^{\infty} \sum_{j=-n}^{n} d_n \frac{\hat{\theta}_{n,j}}{\lambda_n} \psi_{n,j}(x). \quad (9)$$

4 Recovering image rotations

Let images $\xi^{(1)}, \xi^{(2)}$ be related by a rotation $R_* \in SO(3)$, i.e. $\xi^{(2)}(x) = \xi^{(1)}(R_*^{-1}x) \, \forall x \in S^2$. The spherical harmonic coefficients of $\xi^{(1)}, \xi^{(2)}$ satisfy

$$\hat{\xi}^{(2)}_{n,j} = \sum_{l=-n}^{n} \left(\pi_n(R_*)\right)_{j,l} \hat{\xi}^{(1)}_{n,l} \quad |j| \leq n, \; n \in \mathbb{Z}_{\geq 0}. \tag{10}$$

where $\pi_n(R_*)$ is the $d_n \times d_n$ representation matrix corresponding to rotation R_* [2]. To determine R_*, a least squares approach is adopted, i.e. a weighted sum of squares of the difference between the left and right sides of (10) is minimized, with the sum being taken over the appropriate range of j, n. This minimization, in view of the Parseval relation, is equivalent to minimizing $\int_{S^2} \left|\xi^{(2)}(x) - \xi^{(1)}(R^{-1}x)\right|^2 dx$ over $R \in SO(3)$.

The trivial $n = 0$ equation states $\int \xi^{(2)}(x)dx = \int \xi^{(1)}(x)dx$, where the information about the rotation R_* is lost by averaging. For $n = 1$, the rotation is represented by matrix $R_* \in SO(3)$ and $\pi_1(R_*) = TR_*T^{-1}$ for some nonsingular matrix T independent of R_*. Then from equation (10),

$$R_*T^{-1} \begin{bmatrix} \hat{\xi}^{(1)}_{1,-1} \\ \hat{\xi}^{(1)}_{1,0} \\ \hat{\xi}^{(1)}_{1,1} \end{bmatrix} = T^{-1} \begin{bmatrix} \hat{\xi}^{(2)}_{1,-1} \\ \hat{\xi}^{(2)}_{1,0} \\ \hat{\xi}^{(2)}_{1,1} \end{bmatrix} = \mathbf{u}, \tag{11}$$

which may be solved for R_* (Ch.9 [2]). However, the solution is not unique: If R_0 is any such rotation that solves (11) then so do all rotations of the form $R = R_0 \cdot R_1$ where R_1 is a rotation that leaves $\mathbf{u}$ in (11) fixed. Let $\hat{\mathbf{n}}$ denote the unit matrix along $\mathbf{u}$. Then R_1 is of the form $R_1 = R_{\hat{\mathbf{n}}}(\theta)$. And if R_2 is the rotation that rotates the unit vector along z-axis to $\hat{\mathbf{n}}$, then $R_1 = R_{\hat{\mathbf{n}}}(\theta) = R_2 \cdot R_z(\theta) \cdot R_2^{-1}$.

All that remains now is to determine the angle θ. This is done by evaluating (10) at R_* for $n > 1$,

$$\pi_n(R_z(\theta))\pi_n^*(R_2) \begin{bmatrix} \hat{\xi}^{(1)}_{1,-1} \\ \hat{\xi}^{(1)}_{1,0} \\ \hat{\xi}^{(1)}_{1,1} \end{bmatrix} = \pi_n^*(R_0 {\cdot} R_2) \begin{bmatrix} \hat{\xi}^{(2)}_{1,-1} \\ \hat{\xi}^{(2)}_{1,0} \\ \hat{\xi}^{(2)}_{1,1} \end{bmatrix}.$$

But $\pi_n(R_z(\theta))$ is simply the diagonal matrix $diag[e^{-in\theta}, e^{-i(n-1)\theta}, \ldots, e^{i(n-1)\theta}, e^{in\theta}]$. Thus, in component form the set of equations yields

$$e^{in\theta} \, \hat{\xi}^{(1)}_{n,j} \; = \; \hat{\xi}^{(2)}_{n,j}, \quad |j| \leq n \tag{12}$$

any of which can be solved to determine θ.

5 Experimental Results

This section presents experimental results for visual homing using the catadioptric images. The images were captured by a camera (mounted on a robot) pointed upwards at a hyperbolic mirror. The image database is publicly available online [1] and was generated using an ImagingSource DFK 4303 camera, an ActivMedia Pioneer 3-DX robot and a large wide-view hyperbolic mirror. The mobile robot and the imaging system are discussed in detail in [8]. The "original" database consists of 170 omnidirectional images captured at regular grid points on the floor plan as shown in Figure 2.

The tracking and localization performance is discussed as robot navigates from a random starting grid location to an unknown grid location (referred to as 'goal') at which the omnidirectional snapshot is given. The snapshots at current location and grid points neighbouring the current position comprise the templates for matched filtering. Note that in a more general case, the knowledge of the fundamental map T, that relates f and ξ, will allow generation of much larger collection of templates. However, in our experiments with the Bielefeld database [1] in this paper, the templates are restricted to be images at the neighbouring grid locations. For instance, in Figure 2, the shaded box covers all eight neighbours of the grid location $(7, 7)$. The robot successively moves to the grid position corresponding to the template that matches the best.

The translation parameters to be found are the displacement d and the planar direction vector $\hat{\mathbf{u}}$. The rotation parameter to be determined is the azimuth angle ϕ_* since only the rotation (of robot) about the vertical axis is allowed. The online database captures the scene at each grid point for a fixed angular orientation of the robot; thus random azimuth rotations were introduced for the purpose of performance evaluation. The relations in (9) and (12) are computed for all $n \leq B$, for some positive integer B, also referred to as the *bandwidth* of spherical Fourier transform (see Ch. 9 [2]).

The trajectory taken by the robot as it travels from $(0, 0)$ to $(7, 12)$ is shown with solid line in Figure 2. The omni-directional scenes captured by the robot at marked points (shaded dots) along the trajectory are shown in Fig 1. The translation from origin to the goal corresponds to $d_* = (\sqrt{12^2 + 7^2}) \cdot \Delta \approx 13.89 \, \Delta$ and $\hat{\mathbf{u}}_* = \frac{12\Delta}{d_*}\hat{\mathbf{u}}_*^{(x)} + \frac{7\Delta}{d_*}\hat{\mathbf{u}}_*^{(y)} \approx (.86, .50)$, respectively, where $\Delta = 30cm$ is the smallest physical distance between two grid points. The true azimuth angle between the pose at $(0, 0)$ and $(7, 12)$ is $\phi_* = 31°$. The estimates of the unit direction vector and the rotation angle are tabulated in Table 1 for various bandwidths. The error in estimating the rotation angle is the absolute dif-

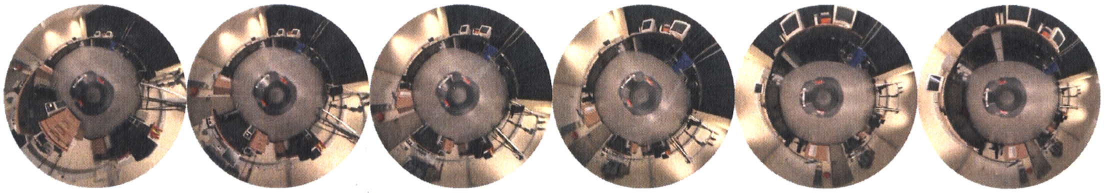

Figure 1. Omni-directional scenes captured at various points along the trajectory shown in solid line in Fig 2.

Bandwidth B	Estimate of $(\hat{\mathbf{u}}_*, d_*)$ $(\hat{\mathbf{u}}_x, \hat{\mathbf{u}}_x, \hat{d})$	Estimation error $\lvert\lvert d\hat{\mathbf{u}} - d_*\hat{\mathbf{u}}_*\rvert\rvert_2$	Estimate of θ_* ϕ	Estimation error $\lvert\hat{\phi} - \phi_*\rvert^2$
6	$(0.92, 0.40, 15.23\Delta)$	$\sqrt{5}\Delta$	60°	29°
12	$(0.91, 0.42, 14.32\Delta)$	$\sqrt{2}\Delta$	30°	1°
18	$(0.86, 0.50, 13.89\Delta)$	0	30°	1°

Table 1. Tracking performance at various bandwidths.

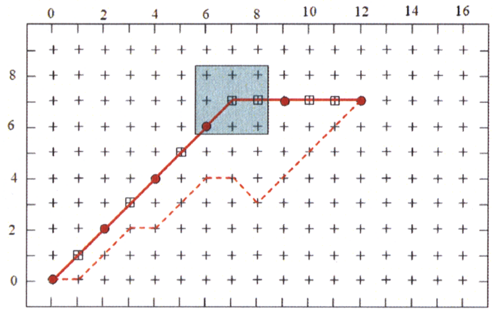

Figure 2. Grid field of robot's environment [8].

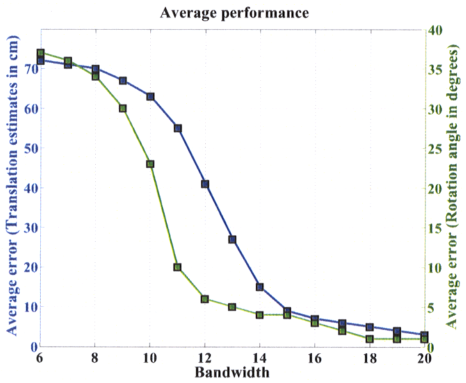

Figure 3. MSE (averaged over trajectories in grid-field of Fig 2) plotted against bandwidth.

ference between the true and the computed rotation and the error in determining the translation is measured as the Euclidean distance $||d\hat{\mathbf{u}} - d_*\hat{\mathbf{u}}_*||_2$, between the true and the estimated destination points. The robot traces its path back to $(0, 0)$ from $(7, 12)$ along the trajectory shown with "dashed" lines in Figure 2. The average estimation accuracy for the homing algorithm is plotted in Fig 3. The mean squared error along random trajectories in the grid field of the robot is plotted against the bandwidth.

References

[1] Panoramic image database, http://www.ti.uni-bielefeld.de/html/research/avardy/index.html.

[2] G. Chirikjian and A. Kyatkin. *Engg. Applications of Non-commutative Harmonic Analysis: With Emphasis on Rotation & Motion Groups*. CRC, 2000.

[3] S. Li. Full-view spherical image camera. In *Int. Conf. Pattern Recognition*, pages 386–390, 2006.

[4] R. Orghidan, E. M. Mouaddib, and J. Salvi. Omnidirectional depth computation from a single image. In *Int. Conf. Robotics Automation*, Apr 2005.

[5] R. Orghidan, J. Salvi, and E. M. Mouaddib. Accuracy estimation of a new omnidirectional 3D vision sensor. In *Int. Conf. Image Proc.*, pages 365–368, Sep 2005.

[6] C. Pegard and E. M. Mouaddib. A mobile robot using a panoramic view. In *ICRA*, pages 89–94, Apr 1996.

[7] T. Svoboda and T. Pajdla. Panoramic cameras for 3D computation. In *Proc. of the Czech Pattern Recognition Workshop*, pages 63–70, Feb 2000.

[8] A. Vardy and R. Moller. Biologically plausible visual homing methods based on optical flow techniques. *Connection Science*, 17:47–89, March 2005.

[9] Y. Yagi. Omni-directional sensing and its applications. *IEICE Trans. Info. systems*, Mar 1999.