

Department of Computer & Information Science

Technical Reports (CIS)

University of Pennsylvania

Year 2004

Motion estimation using a spherical
camera

Ameesh A. Makadia*

Kostas Daniilidis†

*University of Pennsylvania, makadia@seas.upenn.edu

†University of Pennsylvania, kostas@cis.upenn.edu

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-04-10.

This paper is posted at ScholarlyCommons.

http://repository.upenn.edu/cis_reports/3

Motion Estimation Using a Spherical Camera

Ameesh Makadia and Kostas Daniilidis

GRASP Laboratory, University of Pennsylvania, Philadelphia, PA 19104
{makadia, kostas}@cis.upenn.edu

Abstract

Robotic navigation algorithms increasingly make use of the panoramic field of view provided by omnidirectional images to assist with localization tasks. Since the images taken by a particular class of omnidirectional sensors can be mapped to the sphere, the problem of attitude estimation arising from 3D motions of the camera can be treated as a problem of estimating the camera motion between spherical images. This problem has traditionally been solved by tracking points or features between images. However, there are many natural scenes where the features cannot be tracked with confidence. We present an algorithm that uses image features to estimate ego-motion without explicitly searching for correspondences. We formulate the problem as a correlation of functions defined on the product of spheres $S^2 \times S^2$ which are acted upon by elements of the direct product group $SO(3) \times SO(3)$. We efficiently compute this correlation and obtain our solution using the spectral information of functions in $S^2 \times S^2$.

1 Introduction

Estimating the motion of a camera (ego-motion estimation) is a problem that has numerous applications, ranging from mobile robot localization to stereo algorithms. When the motion between frames is large, differential algorithms using optical flow are bypassed in favor of techniques which track features or points between images. Sophisticated feature extractors [9, 5] are often application or scene-dependent in that many parameters must be tuned in order to obtain satisfactory results for a particular data set. Although the tracking of features is considered a familiar and well-understood problem, there are many scenes and objects with repeated textures for which features cannot be successfully matched. However, due to the geometry of spherical perspective, a global image transformation which models the general rigid motion of a camera does not exist, and so we cannot altogether abandon the calculation of localized image characteristics. Roy and Cox [8] have treated this approach by computing a cost function based on the variance of point intensities relative to their Euclidean distance. Geyer [2] proposed a 6D Radon transform on the space of Essential matrices parameterized by ordered pairs in the rotation group $SO(3)$. In contrast, we propose an algorithm which circumvents the pitfalls of feature tracking by processing the features directly without searching for the best matches. Our approach culminates with a five-dimensional search for the parameters of motion via an integral transform similar to the Radon in concept.

2 Motion estimation via a Radon transform

We will first introduce the traditional Radon transform as it applies to identifying lines in planar images before we illustrate how we can use similar intuition to identify the correct ego-motion parameters of a spherical camera in motion. The Radon transform will convert a function from data space into parameter space, and for identifying lines on a planar image it is given as

$$G(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(\rho - \langle (x, y), (\cos \theta, \sin \theta) \rangle) dx dy$$

Here $g(x, y)$ is the weighting function, in this case an intensity image. δ is a soft characteristic function, which measures how close the point (x, y) lies to the line defined by (ρ, θ) . Conceptually, for any line (ρ, θ) , $G(\rho, \theta)$ counts

the number of image points which belong to the line given by $\rho - x \cos \theta - y \sin \theta = 0$, weighted by the intensity of the image points. The rigid motion of a camera is given by the pair (R, t) , where $R \in SO(3)$ is the rotation of the camera and t is the translation. We parameterize $SO(3)$ with ZYZ Euler angles $(R(\alpha, \beta, \gamma) = R_z(\gamma)R_y(\beta)R_z(\alpha))$, and since the translation can only be recovered up to scale, we restrict t to be a unit vector, concerned only by the direction of translation. As the Radon transform identifies lines in planar images, we would like to formulate a conceptually similar transform that will identify the five parameters describing the motion of a camera between two image locations I_1, I_2 . For every possible rigid motion given by (R, t) , we want to count the number of point pairs (p_1, p_2) , where $p_1 \in I_1, p_2 \in I_2$, such that (p_1, p_2) satisfies the motion constraint given by $(Rp_1 \times p_2)^T t = 0$, weighted by the similarity of the points p_1, p_2 . This formulation will be robust only if we can find a similarity measure which will identify point pairs only if the points under comparison are projections of the same scene point. With this objective a simple image-based measure will not suffice. Our proposal is to use image features which compute more distinguishing characteristics such as local gradient orientation distributions. A similarity between such features shows greater contrast between image locations that do not represent the same projected scene point [6]. Using this idea of a similarity between features, we can formulate an integral transform to compute the validity of each possible rigid motion:

$$G(R, t) = \int_{p_1 \in S^2} \int_{p_2 \in S^2} g(p_1, p_2) \delta((Rp_1 \times p_2)^T t) dp_1 dp_2$$

Here our soft characteristic function δ measures how close the pair of feature locations p_1 and p_2 come to satisfying the motion constraint given by $(Rp_1 \times p_2)^T t = 0$. Our weighting function $g(p_1, p_2)$ measures the similarity between features located at points p_1 and p_2 , and is given as

$$g(p_1, p_2) = \{e^{-\|p_1 - p_2\|} \text{ if features have been extracted at } p_1 \text{ and } p_2, 0 \text{ otherwise}\},$$

where $\|p_1 - p_2\|$ is a measure of the difference between two features. Notice that the domain both our weighting function and characteristic function is the manifold $S^2 \times S^2$, since (p_1, p_2) is an ordered pair of points on the sphere S^2 . Since we have restricted t to be a unit vector, we can write $t = R_z R_y e_3$, where $(R_z R_y) \in SO(3)$. Consequently, points in our parameter space can be identified with elements of the direct product group $SO(3) \times SO(3)$. Thus, the functions g, δ are defined on the homogeneous space $S^2 \times S^2$ of the group $SO(3) \times SO(3)$, of which elements of our parameter space belong. In the following section we will show how to utilize this group theoretic framework to compute $G(R, t)$ using the harmonic analysis of functions defined on the space $S^2 \times S^2$.

3 Motion estimation as correlation

If we substitute $t = R_z R_y e_3 = R_2 e_3$ into the characteristic function δ , our integral transform becomes

$$G(R_1, R_2) = \int_{p_1 \in S^2} \int_{p_2 \in S^2} g(p_1, p_2) \delta((R_1 p_1 \times p_2)^T R_2 e_3) dp_1 dp_2 \quad (1)$$

$$= \int_{p_1 \in S^2} \int_{p_2 \in S^2} g(p_1, p_2) \delta((R_1^T R_1 p_1 \times R_2^T p_2)^T e_3) dp_1 dp_2 \quad (2)$$

$$G(R_3, R_2) = \int_{p_1 \in S^2} \int_{p_2 \in S^2} g(p_1, p_2) \delta((R_3^T p_1 \times R_2^T p_2)^T e_3) dp_1 dp_2, R_3 = R_1^T R_2 \quad (3)$$

$G(R_3, R_2)$ is now a correlation of functions defined on the product of spheres $S^2 \times S^2$. The correlation *shift* in this case is performed by elements of the group $SO(3) \times SO(3)$. As explained in detail in [7, 4], a correlation between functions defined on the sphere S^2 , where the shift is given by an element of $SO(3)$, can be computed efficiently using a Spherical Fourier Transform (SFT). We now proceed to extend this development to consider the direct product group $SO(3) \times SO(3)$, beginning with a short exposition of spherical harmonic analysis. Readers are referred to [1] for extensive information regarding the computation of a discrete SFT.

As the angular portion of the solution to Laplace's equation in spherical coordinates, the spherical harmonic functions Y_m^l form a complete orthonormal basis over the unit sphere:

$$Y_m^l(\theta, \phi) = (-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_m^l(\cos \theta) e^{im\phi}$$

where $P_m^l(\cos(\theta))$ are associated Legendre polynomials. Thus, for any function $f(\omega) \in L^2(S^2)$, we have a Spherical Fourier Transform (SFT) given as

$$f(\omega) = \sum_{l \in \mathbb{N}} \sum_{|m| \leq l} \hat{f}_m^l Y_m^l(\omega) \quad (4)$$

$$\hat{f}_m^l = \int_{\omega \in S^2} f(\omega) \overline{Y_m^l(\omega)} d\omega \quad (5)$$

An important property of the spherical harmonic functions Y_m^l is

$$Y_m^l(R^{-1}\eta) = \sum_{|k| \leq l} Y_k^l(\eta) U_{km}^l(R), \quad (6)$$

where the $(2l+1) \times (2l+1)$ matrices U^l are the irreducible unitary matrix representations of the transformation group $SO(3)$, whose elements are given by

$$U_{mk}^l(R) = e^{-im\alpha} P_{mk}^l(\cos \beta) e^{-ik\gamma}. \quad (7)$$

The P_{mk}^l are the generalized Legendre polynomials. From (6) we obtain a Shift Theorem relating coefficients of rotated functions makadia03cvpr:

$$h(\omega) = f(R^{-1}\omega) \Leftrightarrow \hat{h}_m^l = \sum_{|k| \leq l} \hat{f}_k^l U_{mk}^l(R) \quad (8)$$

This Shift Theorem (8) shows us that the U^l matrix representations of the rotation group $SO(3)$ are the spectral analogue to 3D rotations. As vectors in \mathbb{R}^3 are rotated by orthogonal matrices under rotation, the $(2l+1)$ -length complex vectors \hat{f}^l , comprised of all coefficients of degree l , are transformed by the unitary matrices U^l .

As expected, this theory extends to the direct product group $SO(3) \times SO(3)$ acting on the homogenous space $S^2 \times S^2$. The expansion for functions in $S^2 \times S^2$ is given as

$$f(\omega_1, \omega_2) = \sum_{l \in \mathbb{N}} \sum_{|m| \leq l} \sum_{n \in \mathbb{N}} \sum_{|p| \leq n} \hat{f}_{mp}^{ln} Y_m^l(\omega_1) Y_p^n(\omega_2) \quad (9)$$

$$\hat{f}_{mp}^{ln} = \int_{\omega_1 \in S^2} \int_{\omega_2 \in S^2} f(\omega_1, \omega_2) \overline{Y_m^l(\omega_1)} \overline{Y_p^n(\omega_2)} d\omega_1 d\omega_2 \quad (10)$$

A Shift theorem also exists for functions on $S^2 \times S^2$:

$$h(\omega_1, \omega_2) = f(R_1^T \omega_1, R_2^T \omega_2) \Leftrightarrow \hat{h}_{mp}^{ln} = \sum_{|r| \leq l} \sum_{|q| \leq n} U_{rm}^l(R_1) U_{qp}^n(R_2) \hat{f}_{rq}^{ln} \quad (11)$$

We will now use these results to show how to compute the correlation efficiently.

3.1 Algorithm

Expanding our correlation (3) with (11,10), we get:

$$G(R_3, R_2) = \int_{p_1 \in S^2} \int_{p_2 \in S^2} \left[\sum_s \sum_{|t| \leq s} \sum_u \sum_{|v| \leq u} \hat{g}_{tv}^{su} Y_v^u(p_2) Y_t^s(p_1) \right] \quad (12)$$

$$\left[\sum_l \sum_{|m| \leq l} \sum_n \sum_{|p| \leq n} \sum_{|r| \leq l} \sum_{|q| \leq n} \overline{U_{rm}^l(R_3) U_{qp}^n(R_2)} \overline{\hat{\delta}_{rq}^{ln} Y_p^n(p_2) Y_m^l(p_1)} \right] dp_1 dp_2 \quad (13)$$

From the orthogonality of the spherical harmonics this reduces to

$$G(R_3, R_2) = \sum_l \sum_{|m| \leq l} \sum_n \sum_{|p| \leq n} \sum_{|r| \leq l} \sum_{|q| \leq n} \overline{U_{rm}^l(R_3) U_{qp}^n(R_2)} \hat{f}_{mp}^{ln} \overline{\hat{\delta}_{rq}^{ln}} \quad (14)$$

From the homomorphism property of the representations U , we know that

$$U_{mn}^l(R(\alpha, \beta, \gamma)) = \sum_{|k| \leq l} e^{-im(\gamma + \frac{\pi}{2})} e^{-ik(\beta + \pi)} e^{-in(\alpha + \frac{\pi}{2})} P_{mk}^l(0) P_{kn}^l(0)$$

Using this to expand (14), and by defining $R_3 = R_z(\gamma - \frac{\pi}{2})R_y(\beta - \pi)R_z(\alpha + \frac{\pi}{2})$ and $R_2 = R_z(\eta - \frac{\pi}{2})R_y(\xi - \pi)$, we get

$$G(R_3, R_2) = \sum_{lmrk} \sum_{npqj} P_{rk}^l(0) P_{km}^l(0) P_{qj}^n(0) P_{jp}^n(0) e^{i(k\beta + r\alpha + m\gamma + j\xi + p\eta + q\frac{\pi}{2})} \hat{g}_{np}^{ln} \overline{\hat{\delta}_{rq}^{ln}} \quad (15)$$

As it happens, the exponentials in $G(R_3, R_2)$ are orthogonal to the Fourier basis for the circle, so in fact we can take a 5-D Fourier transform of G and obtain

$$\hat{G}_{rkmjp} = \sum_l \sum_n \sum_{|q| \leq n} P_{rk}^l(0) P_{km}^l(0) P_{qj}^n(0) P_{jp}^n(0) e^{iq\frac{\pi}{2}} \hat{g}_{np}^{ln} \overline{\hat{\delta}_{rq}^{ln}} \quad (16)$$

Thus, the Fourier coefficients \hat{G} of our correlation (15) can be computed directly from \hat{g} and $\hat{\delta}$. Note also that the resolution of our correlation grid directly depends upon the band-limit we assume for our functions g , δ . If our band-limit is chosen to be L , we will obtain a result that is accurate up to $\pm(180/(2L + 1))^\circ$ for each parameter.

3.2 Refining the estimate

For a reasonable selection of the variable $L = 20$, we will obtain an estimate with $\pm 4.4^\circ$ accuracy. The computational load required to obtain an estimate of sub-degree accuracy is infeasible. However, we can practically compute (3) directly in a window of 8.8° in each parameter to localize our solution. It is important to note that since we are only refining our solution (we assume it is correct, and only wish to localize it), we can use the initial estimate to prune the feature pairs which are deemed outliers, thus greatly reducing the computational load.

4 Experiments

In this section we will present some practical considerations regarding the computation of our correlation function (15) and its coefficients, followed by some experimental results on real data. For comparison, we use a popular feature extractor/tracker to generate correspondences from which we estimate the motion. For a ground truth result, we track by hand 30 image points from which we again estimate the motion.

4.1 Spherical Images

Catadioptric systems with a unique effective viewpoint have been proven to be convex reflective surfaces of revolution with a parabolic or hyperbolic profile. Geyer and Daniilidis [3] showed that such projections are equivalent to a projection on the sphere followed by a projection from a point on the sphere axis to the plane. In the parabolic case, the second projection is a stereographic projection from the sphere to the catadioptric plane (also the image plane):

$$\begin{aligned} u &= \cot \frac{\theta}{2} \cos \phi \\ v &= \cot \frac{\theta}{2} \sin \phi. \end{aligned}$$

Given a calibrated camera and catadioptric image $I(u, v)$ we define its inverse stereographic mapping onto the sphere as

$$I_S(\theta, \phi) \stackrel{def}{=} I\left(\cot \frac{\theta}{2} \cos \phi, \cot \frac{\theta}{2} \sin \phi\right).$$

This mapping allows us to interpolate an image on the sphere given a catadioptric image. The range of this mapping is only limited by the field of view of the original catadioptric system, and so to fully image the sphere a 360° field of view catadioptric system would be required.

4.2 Image acquisition

To obtain spherical images, we used a catadioptric system consisting of a Nikon Coolpix 995 digital camera along with a parabolic mirror attachment produced by Remote Reality. The mirror's field of view is 212° . The size of the original catadioptric images was 2048×1536 pixels without compression, and the parabolic mirror filled up a region of approximately 1400×1400 pixels. The images are mapped to the sphere by interpolating onto the θ - ϕ plane, where angular sampling is uniform. Figure 1 shows a sample catadioptric image obtained from a parabolic mirror and its corresponding projection onto the sphere.



Figure 1: On the left is a parabolic catadioptric image. In the middle is the spherical image represented on the θ - ϕ plane, and on the right is the image displayed on the sphere.

4.3 Feature extraction

To extract features from our catadioptric images we use the Scale Invariant Feature Transform (SIFT). The SIFT feature extraction algorithm identifies distinguishable feature vectors using a scale-space difference-of-gaussians approach. Feature vectors are computed using neighborhood gradient orientation information. A feature generated from SIFT typically has 128 parameters. To compute difference between two feature vectors in \mathbb{R}^{128} , we simply use the Euclidean distance. Figure 2 shows the feature correspondences you would obtain if you matched SIFT features between two images.

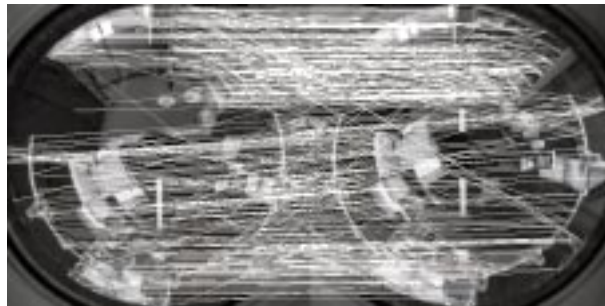


Figure 2: Correspondences generated between two images using SIFT features.



Figure 3: Two catadioptric images taken from a camera moving in only the positive X direction.

We tested our algorithm on the pair of images shown in Figure (3). For comparison, we searched for correspondences between our feature sets, and then applied Levenberg-Marquardt minimization to find the five parameters of motion. To determine the validity of each solution, we hand-tracked 30 features p_i, q_i and computed the error $\sum_{i=1}^{30} |(R_{est} p_i \times q_i)^T t_{est}|^2$. The error of our algorithm after performing a refinement in the solution window using a direct computation was 0.0159, and the error of LM minimization was a comparable 0.0179. To understand the magnitude of these errors, we also performed LM on our hand tracked correspondences, and the error of these matches using its own estimate was 0.0074;

5 Conclusion

In many instances, Fourier based algorithms offer a significant advantage compared to direct, brute-force spatial computations. Global Fourier techniques are generally frowned upon when dealing with motion estimation problems because as global operators they cannot account for signal alterations introduced by occlusion, depth-variations, and a limited field of view. We avoid these pitfalls by analyzing the spectral information of feature-based signals rather than the original spherical images. Our preliminary results indicate that with a reasonable computational load we can obtain a motion estimate up to a small window. If necessary, a fast direct computation will deliver a refined solution.

References

- [1] J. Driscoll and D. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15:202–250, 1994.
- [2] C. Geyer. Euclid meets fourier. In *Workshop on Omnidirectional Vision*, Prague, 2004.
- [3] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 43:223–243, 2001.
- [4] J. A. Kovacs and W. Wriggers. Fast rotational matching. *Biological Crystallography*, 58:1282–1286, 2002.
- [5] D. Lowe. Sift (scale invariant feature transform): Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [6] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. Int. Conf. on Computer Vision*, pages 1150–1157, Kerkyra, Greece, Sep. 20-23, 1999.
- [7] A. Makadia, L. Sorigi, and K. Daniilidis. Rotation estimation from spherical images. In *To Appear in Proc. Int. Conf. on Pattern Recognition*, Cambridge, UK, 2004.
- [8] S. Roy and I. Cox. Motion without structure. In *Proc. Int. Conf. on Pattern Recognition*, Vienna, Austria, 1996.
- [9] J. Shi and C. Tomasi. Good features to track. In *IEEE Conf. Computer Vision and Pattern Recognition*, Hilton Head Island, SC, June 13-15, 1994.