# Planar Ego-Motion Without Correspondences

Ameesh Makadia, Dinkar Gupta, and Kostas Daniilidis *

GRASP Laboratory
Department of Computer and Information Science
University of Pennsylvania, Philadelphia, PA 19104
{makadia, dinkar, kostas}@cis.upenn.edu

## Abstract

*General structure-from-motion methods are not adept at dealing with constrained camera motions, even though such motions greatly simplify vision tasks like mobile robot localization. Typical ego-motion techniques designed for such a purpose require locating feature correspondences between images. However, there are many cases where features cannot be matched robustly. For example, images from panoramic sensors are limited by nonuniform angular sampling, which can complicate the feature matching process under wide baseline motions. In this paper we compute the planar ego-motion of a spherical sensor without correspondences. We propose a generalized Hough transform on the space of planar motions. Our transform directly processes the information contained within all the possible feature pair combinations between two images, thereby circumventing the need to isolate the best corresponding matches. We generate the Hough space in an efficient manner by studying the spectral information contained in images of the feature pairs, and by re-treating our Hough transform as a correlation of such feature pair images.*

## 1 Introduction

Estimating the planar 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 [11, 8] 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 practical scenarios (depending on properties of the imaging sensor, or scenes and objects with repeated textures) for which features cannot be successfully matched. Take for example omnidirectional camera systems, which have become synonymous with mobile robots. The panoramic view which makes such sensors so appealing is also being represented by relatively fewer pixels (per viewing angle). This fact, combined with the projection geometry of such sensors, makes the problem of matching points between images quite difficult under many circumstances. Since a particular class of central omnidirectional sensors can be mapped to the sphere, our problem of planar ego-motion from spherical images encompasses images obtained from such panoramic sensors. However, due to the geometry of the spherical perspective projection, a global image transformation which models rigid motions of a camera does not exist, and so we cannot altogether abandon the calculation of localized image characteristics. Previously, Aloimonos and Hervé [1] showed the rigid motion of planar patches can be estimated without correspondences using a binocular stereo setup. In [2] Antone and Teller use a Hough transform on lines in images to identify vanishing points (which are used in rotational alignment). A subsequent Hough transform on feature pairs in rotationally aligned images is used to estimate the direction of translation. The computational complexity of a direct computation is circumvented by pruning possible feature pairs and only desiring an rough estimate of the solution, which is used to initialize an EM algorithm. Roy and Cox [10] contributed a correspondenceless motion algorithm by statistically modeling the variance of intensity differences between points relative to their Euclidean distance. This model is then used to estimate the likelihood of assumed motions. Geyer et al [5] proposed a 6D Radon transform on the space of Essential matrices parameterized by ordered pairs in the rotation group $SO(3)$. Both proposed solutions address the general ego-motion problem, but in this paper we consider the constrained problem of a camera moving in a plane. We revisit the familiar two-view epipolar constraint and examine

*The authors are grateful for support through the following grants: NSF-IIS-0083209, NSF-IIS-0121293, NSF-EIA-0324977, NSF-CNS-0423891, NSF-IIS-0431070, and ARO/MURI DAAD19-02-1-0383.

what happens when we are able to parameterize planar motions with only rotations. Furthermore, we circumvent the pitfalls of feature tracking by processing image features directly without searching for the best matches. The resulting planar constraint deals with these feature pairs which can alternatively be considered elements of the product of spheres $S^2 \times S^2$. To compensate for the fact that we must consider all possible feature pairs between images, we introduce a generalized Hough transform that implicitly accounts for the fact that the majority of our information does not represent the correct motion. We explore the efficient computation of our Hough space by realizing from the planar epipolar constraint that the elements of the Hough space can be identified with the direct product group $SO(3) \times SO(3)$, which acts transitively on the domain of our feature images. Such a group theoretic framework allows us to compute our Hough space very efficiently.

The organization of this paper is as follows. In section 2 we will introduce the traditional Hough transform on the space of lines and generalize it for the space of planar motions. We will follow this with a brief exposition of harmonic analysis on the sphere and product of spheres, and its application in formulating the generalized Hough as a correlation (Section 3.1). Practical considerations and experimental results follow in Section 4 and we will wrap up our discussion with some concluding remarks in Section 5.

# 2 Motion estimation as Hough

Consider a camera viewing points $P_{1\cdots j}$, which are given as vectors relative to the camera's reference frame. The spherical perspective projection maps world points $P_{1\cdots j}$ to image points $p_{1\cdots j}$, where $p_i = P_i/||P_i||$. Now assume the camera undergoes a rigid motion in the X-Y plane given by the pair $(R, t)$, where $R = R_z(\alpha)$ is a rotation about the Z-axis by $\alpha$ degrees and $t$ is a vector in the X-Y plane. We wish to reflect the fact that without knowledge of the environment, $t$ can only be estimated up to scale, and thus can be defined with only one free parameter. We will write $t = R_z(\theta)\vec{e_1}$. From the camera's new coordinate frame, the world points are now given as $Q_{1\cdots j} = RP_{1\cdots j} + t$, and the world points $Q_{1\cdots j}$ map to image points $q_{1\cdots j}$, where $q_i = Q_i/||Q_i||$. Since $p_i$ and $q_i$ do not retain any information about the depth of the world points $P_i$ and $Q_i$, we can only infer that the $Rp_i$, $q_i$, and $t$ lie on the same plane: $(R_z(\alpha)p_i \times q_i)^T R_z(\theta)\vec{e_1} = 0$. This condition is just the familiar two-view epipolar constraint restricted to camera motions in the X-Y plane. However, we are not necessarily limited to treating motions only in this plane. Consider a camera moving in an arbitrary plane given by the angles $\beta$, and $\gamma$ (the vector $R_z(\gamma)R_y(\beta)\vec{e_3}$ is orthogonal to all vectors lying in the plane). Since $R_z(\gamma)R_y(\beta)$ and $-R_z(\gamma)R_y(\beta)$ describe the same plane, we can restrict $(\beta, \gamma)$ to the hemisphere: $\beta \in [0, \frac{\pi}{2}]$ and $\gamma \in [0, 2\pi)$. Since we can define any plane of motion with rotations, it is immediately clear that motions on the plane $(\beta, \gamma)$ are simply a coordinate frame rotation from motions in the X-Y plane. In other words, if the point pairs $(p_i, q_i)$ satisfy a motion in the plane $(\beta, \gamma)$, then the point pairs $(p'_i, q'_i)$, where $\{p'_i, q'_i\} = R_y(-\beta)R_z(-\gamma)\{p_i, q_i\}$, will satisfy the same motion in the X-Y plane. This epipolar constraint for general planar motions can be expressed as:

$$(R_z(\alpha)(R_z(\gamma)R_y(\beta))^T p_i \times (R_z(\gamma)R_y(\beta))^T q_i)^T R_z(\theta)\vec{e_1} = 0 \tag{1}$$

As always, the translational component of ego-motion can only be recovered up to scale, so in general there are always at least two solutions $(R_z(\theta), R_z(\theta+\pi))$, which satisfy (1). We proceed with a discussion on how we can explore this planar epipolar constraint to obtain an ego-motion solution without finding image correspondences.

## 2.1 Hough on the space of planar motions

We will first introduce the Hough transform as it applies to identifying lines in planar images before we illustrate how we can use similar intuition to identify the correct planar ego-motion parameters of a spherical camera. Specifically, we will examine the continuous case of the Hough transform which is often identified with the Radon transform. 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

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

Here $g(x, y)$ is the weighting function, in this case an intensity or gradient image. $\delta$ is a soft characteristic function, which measures how close the point (x,y) lies to the line defined by $(\rho, \theta)$. Intuitively, for any line $(\rho, \theta)$, $G(\rho, \theta)$ 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. As the Radon transform identifies lines in planar images, we would like to formulate a conceptually similar transform that will identify the four parameters describing the planar motion of a camera between two images $I_1, I_2$.

From (1) we have parameterized our planar motion with just rotations. If we parameterize the proper rotation group $SO(3)$ with ZYZ Euler angles $R(\alpha, \beta, \gamma) = R_z(\gamma)R_y(\beta)R_z(\alpha)$, we can rewrite (1) as

$$((QR)^T p_i \times Q^T q_i)^T \vec{e_1} = 0 \tag{2}$$

where $R = R_z(-\alpha)$ and $Q = R_z(\gamma)R_y(\beta)R_z(\theta)$.

For each planar motion $(R, Q)$, we want to count the number of point pairs ($p$, $q$), where $p \in I_1,\ q \in I_2$, such that ($p$, $q$) satisfies the motion constraint (2), weighted by the similarity of the points $p$, $q$. This formulation will be robust only if we find a measure which consistently identifies projections of the same scene point in each image.

Note that in this case we do not have such a strong constraint on false positives, since the presence of outliers is inherently accounted for with a generalized Hough transform, whereas a best-fit minimization requires strong conditions on the quality and quantity of matches for successful ego-motion estimation. With this objective a simple image-based measure will not suffice. Our proposal is to use SIFT image features [9] which compute distinguishing characteristics such as local gradient orientation distributions. Using this idea of a similarity between features, we can formulate an integral transform to compute the validity of each possible rigid motion:

$$G(R,Q) = \int_p \int_q g(p,q)\delta(((QR)^T p \times Q^T q)^T \vec{e_1})dpdq \quad (3)$$

Here the soft characteristic function $\delta$ measures how close the pair of feature locations $p$ and $q$ come to satisfying the motion constraint (2). The weighting function $g(p,q)$ measures the similarity between features located at points $p$ and $q$, and is given as

$$g(p,q) = \begin{cases} e^{-||p-q||} \text{ if features have been extracted at } p \text{ and } q \\ 0 \text{ otherwise} \end{cases}$$

where $||p-q||$ is a measure of difference between two features. Notice that the domain of both our weighting function and characteristic function is the manifold $S^2 \times S^2$, since $(p,q)$ is an ordered pair of points on the sphere $S^2$. Similarly, points in our parameter space can be identified with elements of the direct product group $SO(3) \times SO(3)$. Thus, the functions $g$, $\delta$ are defined on the homogeneous space $S^2 \times S^2$ of the Lie group $SO(3) \times SO(3)$. In the following section we will utilize this group theoretic framework to compute $G(R,Q)$ using the harmonic analysis of functions defined on the space $S^2 \times S^2$.

# 3 Hough as correlation

Taking a close look at (3), we see that the Hough transform $G(R,Q)$ is in fact 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 with detail in [7], 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 [3] 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

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

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

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

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

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 :

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

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, 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

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

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

with a corresponding Shift theorem (in matrix form):

$$h(\omega_1,\omega_2) = f(R_1^T\omega_1, R_2^T\omega_2) \Leftrightarrow \hat{h}_{mp}^{ln} = U^l(R_1)^T \hat{f}^{ln} U^n(R_2) \quad (9)$$

We will now present an efficient computation of the correlation (3) using these results.

### 3.1 Algorithm

Expanding (3) with (9), and from the orthogonality of the spherical harmonics,

$$G(R,Q) = \sum_{l,n} \sum_{|m,r|\leq l} \sum_{|p,q|\leq n} \overline{e^{-im\alpha} U^l_{rm}(Q) U^n_{qp}(Q)} \hat{f}^{ln}_{mp} \hat{\delta}^{ln}_{rq}$$

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

$$\begin{aligned} U^l_{mn}(R(\alpha,\beta,\gamma)) &= \sum_{|k|\leq l} e^{-i(m(\gamma+\frac{\pi}{2})+k(\beta+\pi)+n(\alpha+\frac{\pi}{2}))} \cdot \\ &\quad \cdot P^l_{mk}(0) P^l_{kn}(0) \end{aligned}$$

Defining $Q' = R_z(\gamma - \frac{\pi}{2}) R_y(\beta - \pi) R_z(\theta + \frac{\pi}{2})$, we get

$$\begin{aligned} G(R,Q') &= \sum_{lmrk} \sum_{npqj} P^l_{rk}(0) P^l_{km}(0) P^n_{qj}(0) P^n_{jp}(0) \cdot \\ &\quad \cdot e^{i(k\beta + r\theta + m\gamma + j\gamma + p\beta + q\theta + m\alpha)} \hat{g}^{ln}_{mp} \hat{\delta}^{ln}_{rq} \end{aligned}$$

As it happens, the exponentials in $G(R,Q)$ are orthogonal to the Fourier basis of the circle, so after taking a 4-D Fourier transform of $G$ we obtain

$$\begin{aligned} \hat{G}_{xyzw} &= \sum_{lmrk} \sum_{npqj} P^l_{rk}(0) P^l_{km}(0) P^n_{qj}(0) P^n_{jp}(0) \cdot \\ &\quad \cdot \delta_{(x,m)} \delta_{(y,k+j)} \delta_{(z,m+p)} \delta_{(w,r+q)} \hat{g}^{ln}_{mp} \hat{\delta}^{ln}_{rq}, \end{aligned}$$

where $\delta_{(a,b)}$ is the Kronecker delta. After simplifying,

$$\begin{aligned} \hat{G}_{xyzw} &= \sum_{lnqj} P^l_{(w-q)(y-j)}(0) P^l_{(y-j)(x)}(0) \cdot \\ &\quad \cdot P^n_{(q)(j)}(0) P^n_{(j)(z-x)}(0) \hat{g}^{ln}_{(x)(z-x)} \overline{\hat{\delta}^{ln}_{(w-q)(q)}} \end{aligned}$$

Looking closely at this formula for $\hat{G}$, we can see that it depends solely on the Legendre polynomials $P^l(0)$, and the coefficient matrices $\hat{g}$, $\hat{\delta}$. Since $P^l(0)$ and $\hat{\delta}$ contain no image information, they can be computed and stored offline.

Furthermore, the first SFT of $\delta$ can be computed analytically since $\delta(p,q) = \delta((p \times q)^T \vec{e_1})$ is just a great circle on the sphere for any fixed $p$. This reflects the fact that under a pure translation along the X axis the point $p$ can only move to another point on the great circle intersecting $p$, $\vec{e_1}$, and $-\vec{e_1}$.

Only $\hat{g}$ needs to be computed for each pair of images under consideration. Although we can obtain our ego-motion estimate directly from the 4D IFFT of $\hat{G}$, generating this coefficient grid for large $L, N$ can be quite expensive. For a faster practical implementation we now present an alternative solution.

As we discussed earlier, if the plane of motion is known, we can transform our points $p, q$ so that the effective plane of motion is the X-Y plane. This means for every plane through the origin we can compute a Hough transform to find the angle of rotation and direction of translation. The global maximum over all planes will represent the correct planar ego-motion. For each plane $R_z(\gamma)R_y(\beta)$, our new reference frame can be obtained by shifting $g(p,q)$ : $g(R_z(\gamma)R_y(\beta)p, R_z(\gamma)R_y(\beta)Tq)$. Practically, instead of rotating and interpolating functions on $S^2 \times S^2$, we can achieve the same effect from (9). Searching for motion only in the X-Y plane, our epipolar constraint is

$$(R_z(\alpha - \theta)p_i \times R_z(-\theta)q_i)^T \vec{e_1} = 0 \tag{10}$$

Proceeding with the same exposition as for the case of general planar motion, we find that our correlation function $G$, which is now a function of two rotations about the Z-axis, can be written directly as

$$G(R_z(\theta - \alpha), R_z(\theta)) = \sum_{lnmp} \hat{g}^{ln}_{mp} \hat{\delta}^{ln}_{mp} e^{i(m(\alpha-\theta)-p\theta)}, \tag{11}$$

and the Fourier coefficients of $\hat{G}$ are simply

$$\hat{G}_{k_1 k_2} = \sum_{ln} \hat{g}^{ln}_{k_1 k_2} \hat{\delta}^{ln}_{k_1 k_2} \tag{12}$$

Thus, the Fourier coefficients $\hat{G}$ 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 $g$ and $\delta$. 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 Searching for the plane

We have proposed a fast Hough transform for each plane of motion in our discretized parameter space. Since we identify planes with the angles $\beta \in [0, \frac{\pi}{2}]$, $\gamma \in [0, 2\pi)$, searching through the space of planes is equivalent to searching on the hemisphere. There have been numerous attempts to generate an equidistant (distance meaning geodesic distance) distribution of points on the sphere, and here we adopt a method based on the subdivision of the icosahedron [6]. This allows us to use a fast multi-resolution approach to localize the plane of motion.

## 4 Experiments

In this section we will address some practical considerations regarding the computation of our planar ego-motion parameters, followed by experimental results. For comparison we estimated the Essential Matrix with manually selected correspondences, and also with RANSAC using the set of possible feature pairs as input. The rotation and translation were then decoupled.

### 4.1 Spherical Images

The projections of catadioptric systems with a unique effective viewpoint are equivalent to a projection on the sphere followed by a projection from a point on the sphere axis to the plane [4]. If calibrated, such a sensor enables us to interpolate images on the sphere reflecting a correct spherical perspective projection. To obtain spherical images, we used a system consisting of a Canon Powershot G2 digital camera along with a parabolic mirror attachment produced by Remote Reality. The mirror's field of view is 212°. The images are mapped to the sphere by interpolating onto the $\theta$-$\phi$ grid, where angular sampling is uniform. Figure (1) shows a sample catadioptric image obtained from a parabolic mirror and its corresponding projection onto the sphere.

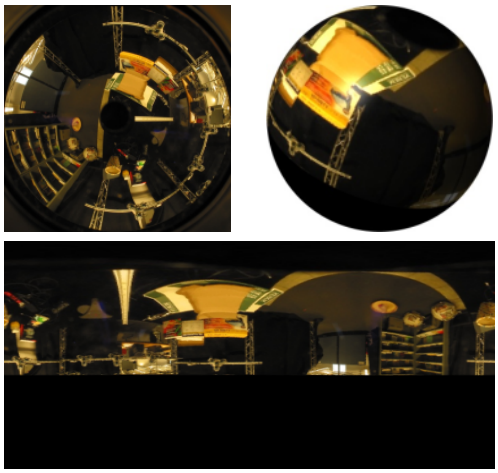

Figure 1: Top left: a parabolic catadioptric image. Bottom: corresponding spherical image on a uniformly sampled $\theta$-$\phi$ grid. Top right: the image as it would appear on the surface of the sphere

We begin with Figure (2), which shows on top a representative pair of spherical images obtained from our catadioptric system. In this experiment the plane of motion is known to be the X-Y plane, so we only need to estimate the rotation angle and translation direction for each image pair. Table (1) shows the results of our estimation for a number of image pairs compared to the estimates obtained from two comparative techniques. Our coefficient grid $\hat{G}$ was computed for a maximum degree $L = N = 32$, so our angular samples were spaced approximately $5.54°$ apart, giving us a solution accurate up to $\pm 2.77°$. In Figure (2), bottom, we show the final discretized Hough space for the roughly translational motion along the X axis. From the grids it is easy to see that the correct solutions are identified. In all examples where the plane of motion is known, our algorithm obtained a solution sufficiently close to the one generated by manually-selected correspondences. Since SIFT detected on the order of 1000 features in the images we used, there

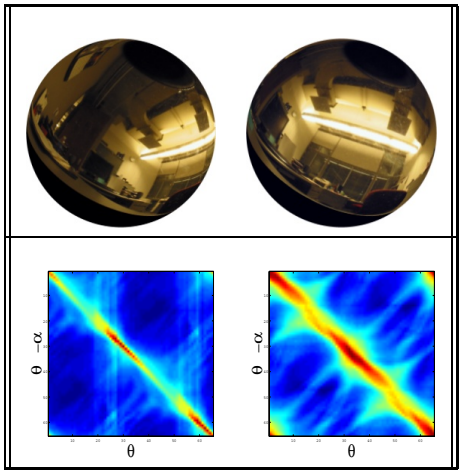


Figure 2: Top: Two spherical images from a camera moving in the X-Y plane. Bottom: The Hough space of two pairs of images where the motion between frames consists of only a translation along the X direction approx. You can see clearly that the two correct solutions at $\theta = 0$ and $\theta = \pi$ have the highest values in the Hough space.

| | | Generalized Hough | RANSAC | Manual Corresp. |
|---|---|---|---|---|
| **Pair 1** | $\alpha$ | $44.3°$ | $33.9°$ | $47.1°$ |
| | $\theta$ | $49.8°$ | $25.1°$ | $44.7°$ |
| **Pair 2** | $\alpha$ | $38.8°$ | $46.2°$ | $35.6°$ |
| | $\theta$ | $27.7°$ | $16.5°$ | $25.0°$ |
| **Pair 3** | $\alpha$ | $0.0°$ | $-4.3°$ | $1.2°$ |
| | $\theta$ | $5.5°$ | $-4.4°$ | $1.6°$ |

Table 1: Estimates of planar motion parameters when the plane of motion is known. In this case only the angle of rotation ($\alpha$) and the direction of translation ($\theta$) need to be estimated. The Hough was computed with $L = N = 32$, giving us an accuracy of $\pm 2.8°$.

were on the order of $10^6$ features pairs, of which $99.9\%$ are guaranteed outliers. Since RANSAC estimates are done with a minimal number of feature pairs, the overwhelming number of outliers drastically effects the average time to find a suitable solution. Also, this algorithm is very susceptible to the accumulation threshold since we don't search for a global maximum but rather terminate once we find the first suitable solution. Even after pruning the feature pairs by eliminating the worst matches, this sensitivity is reflected in the results.

For motion in an unknown plane, visual results are shown in Figure (3), where the Hough space for the rotation angle and translation direction obtained at the final plane. This plane, as described in Section (3.2), was obtained with a coarse-to-fine search through the space of planes done in two iterations. The figure shows the Hough space at the fi-

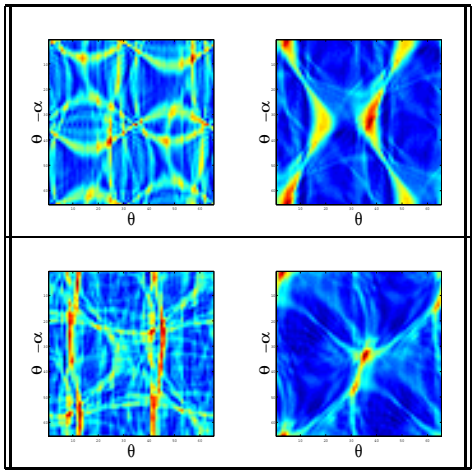


Figure 3: Discretized Hough space for two pairs of test images. On the left is the Hough space for the initial plane (X-Y) and on the right the grid at the plane of the solution. Note the sharp identifiable peak when the solution is found.

| | | Generalized Hough | Manual Correspondences |
|---|---|---|---|
| **Pair 1** | $\beta$ | $173.8°$ | $174.7°$ |
| | $\gamma$ | $135.0°$ | $132.0°$ |
| | $\alpha$ | $44.3°$ | $49.6°$ |
| | $\theta$ | $49.8°$ | $44.3°$ |
| **Pair 2** | $\beta$ | $86.9°$ | $91.0°$ |
| | $\gamma$ | $105.5°$ | $100.6°$ |
| | $\alpha$ | $16.6°$ | $172.4°$ |
| | $\theta$ | $0.0°$ | $1.0°$ |

Table 2: Estimates of planar motion with unknown plane. All four parameters of motion are estimated. The angle of rotation ($\alpha$) and the direction of translation ($\theta$), and the plane of motion ($\beta, \gamma$). For the Hough, the computation was at $L = N = 32$, resulting in an accuracy of $\pm 2.8°$.

nal solution adjacent to the inital starting point (which corresponds to X-Y as the plane of motion). Numerical results of are shown in Table (2). For Pair 1, we visually documented a motion of the camera in the X-Z plane and this was reflected in an angular measurement in $\beta$ of $173.8°$ ($180°$ would be exactly on the X-Z plane). The maximum value in the Hough space is identified with a smooth and strong peak, which indicates that spectral interpolation may be effective in increasing the resolution and accuracy of our result.

The computational time to compute (12) for $L = N = 32$ is $0.2$ sec on a typical 2GHz machine, and obtaining the Hough space via a 2D IFFT requires $1$ ms. If we are performing the test for a plane hypothesis, the rotation of the function $g(p, q)$ takes an additional $1.57$ sec. In total, the time it takes to generate our Hough space for any plane is $1.77$ sec. All of these times were generated in Matlab.

# 5 Conclusion

Fourier techniques conventionally attempt to perform a computation on the global spectral content of an image. In problems of matching, for example, one standard approach is to correlate the spectrum of the target and pattern images. These approaches are limited because as global operators they cannot account for signal alterations introduced by occlusion, depth-variations, and a limited field of view. Instead of trying to estimate motion using the spectral components of the intensity images, we perform our Fourier decomposition on the feature information stored within the images. The epipolar geometric constraints provided by planar ego-motion allow us to establish a generalized Hough transform which falls between a pure spectral computation (global) and a best-fit correspondence technique (local). Results indicate that planar ego-motion can be correctly computed for a spherical camera using only image features and without first finding image correspondences.

# References

[1] J. Aloimonos and J. Y. Hervé. Correspondenceless stereo and motion: Planar surfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12:504–510, 1990.

[2] M. Antone and S. Teller. Scalable, extrinsic calibration of omni-directional image networks. *International Journal of Computer Vision*, 49:143–174, 2002.

[3] J. Driscoll and D. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15:202–250, 1994.

[4] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 43:223–243, 2001.

[5] C. Geyer, S. Sastry, and R. Bajcsy. Euclid meets fourier. In *Workshop on Omnidirectional Vision*, Prague, 2004.

[6] L. Kobbelt. $\sqrt{3}$ subdivision. In *Proceedings of ACM SIGGRAPH*, pages 103–112, 2000.

[7] J. A. Kovacs and W. Wriggers. Fast rotational matching. *Biological Crystallography*, 58:1282–1286, 2002.

[8] D. Lowe. Sift (scale invariant feature transform): Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[9] 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.

[10] S. Roy and I. Cox. Motion without structure. In *Proc. Int. Conf. on Pattern Recognition*, Vienna, Austria, 1996.

[11] 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.