Rotation Recovery from Spherical Images without Correspondences

Ameesh Makadia and Kostas Daniilidis *
GRASP Laboratory, University of Pennsylvania, Philadelphia, PA 19104
{makadia, kostas}@cis.upenn.edu

Abstract

This paper addresses the problem of rotation estimation directly from images defined on the sphere and without correspondence. The method is particularly useful for the alignment of large rotations and has potential impact on 3D shape alignment. The foundation of the method lies in the fact that the spherical harmonic coefficients undergo a unitary mapping when the original image is rotated. The correlation between two images is a function of rotations and we show that it has an SO(3)-Fourier transform equal to the pointwise product of spherical harmonic coefficients of the original images. The resolution of the rotation space depends on the bandwidth we choose for the harmonic expansion and the rotation estimate is found through a direct search in this 3D discretized space. A refinement of the rotation estimate can be obtained from the conservation of harmonic coefficients in the rotational shift theorem. A novel decoupling of the shift theorem with respect to the Euler angles is presented and exploited in an iterative scheme to refine the initial rotation estimates. Experiments show the suitability of the method for large rotations and the dependence of the method on bandwidth and the choice of the spherical harmonic coefficients.

*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.

1. Introduction

In this paper, we introduce a new aspect to the global estimation of motion directly from intensity images without features and without correspondence or optical flow. We study the case of three-dimensional rotations on images defined on spheres, a domain where any perspective projection can be mapped. Research in localization or estimation of camera displacement has been dominated by the detection and tracking of landmarks or salient features in images. In the vast majority of discrete motion approaches, the features are points. Only in the case of differential motion, appearing in the image as optical flow, direct approaches have been employed using spatiotemporal image derivatives only. The latter approaches have been quite successful in the case of affine or projective transformations in significant areas of images and particularly in the production of mosaics. Though designed for image velocities, their hierarchical implementations can also handle larger motions.

Our motivation comes from two directions: the first one is technological, namely, the availability of wide field of view cameras which enable a persistence of image content. The second motivation comes from biological findings that motion perception and navigation are based on global matched filters implemented by neurons of almost spherical receptive fields.

We represent images defined on the sphere with the spherical harmonic expansion. We recall from group theory that the natural group action on the sphere is a three-dimensional rotation and from harmonic analysis that spherical harmonic coefficients obey a shift theorem, analogous to the Fourier coefficients on the real line. This shift theorem states that the rotation of an image corresponds to a unitary mapping of its spherical harmonic coefficient vectors. This unitary matrix depends on the unknown 3D rotation in a way that does not allow the extraction of the unknown 3D rotation in closed form. Moreover, such coefficients vectors are hardly preserved because of appearing and disappearing content during camera motion. It is rather plausible to use the correlation between the two images in such a case. This correlation is a function of a three-dimensional rotation. Functions defined on SO(3) have fast SO(3)-Fourier transforms (called SOFT [13]) which are closely related to the spherical harmonic expansions. We show that the SOFT of the correlation can be computed with the pointwise multiplication of the spherical harmonic coefficients of the two images. Recovering the original correlation function is achieved with a fast inverse SOFT

transform and the actual rotation is obtained by a search in the space of Euler angles. The accuracy of the estimated rotation depends on the resolution of the rotation space which is restricted by the bandwidth, the number of spherical harmonic coefficients we keep from the original images. When the rotation has been constrained to lie in a small resolution-induced interval we can apply a refinement based on the shift theorem or even any algorithm suitable for small rotations. We propose a refinement based on a novel decoupling of the shift theorem and a two-step iteration. We present results for large rotations up to 60 degrees about any axis and we study the dependence of the method on the selection of spherical harmonic coefficients.

Two approaches are the closest to ours: The computation of the correlation function by [19, 14] does not make any use of the SO(3) Fourier transform but is recovered with conventional FFT's in a way described briefly in the correlation section below. The use of the shift theorem in [3] makes use of spherical harmonic coefficients of a particular order to obtain the Euler angles in closed form. We show in the experiments that such a selection suffers from instabilities because there is no guarantee that the particular frequencies contain energy variation sufficient for rotation estimation. The dual problem to rotation estimation, finding rotation invariant signatures, has been a challenging problem in the matching of 3D-shapes starting from the Extended Gaussian Image [9, 10] up to the recent spherical harmonic methods in [11, 12]. Regarding techniques without correspondences, it is basic knowledge that we can compute affine invariants and that we can compute an affine transformation from combinations of image moments [22]. Work has been done on the computation of 3D-rotations from area-based features [15, 24] but without a computational procedure for rotation recovery. Direct estimation of motion from images has recently gained interest in the problem of localization based on reference views. Appearance-based localization [16] has been shown to be successful by applying PCA to a large set of reference views. Pajdla and Hlavac [26] studied the rotation just around the optical axis of panoramic systems and constructed a rotation-invariant representation.

Direct methods in establishing collineations directly from image intensities or their derivatives have existed for a long time [2, 23, 25, 21]. General differential motions of a calibrated camera are estimated in [6]. The disadvantage of existing direct methods is that in the case of larger motions they rely on a series of iterations involving a warping of the image at every step and in principle they are iterative closest point (ICP) algorithms requiring an initialization close to ground truth.

2. Rotation Estimation

If asked about the problem of estimating rotations between spherical images, the outside observer may offer this expensive yet effective algorithm: compare the two images for every possible rotation, and the rotation under which the two images are the most similar is the solution. Given black-box operators for image rotation and comparison, we could construct a function to determine the strength of each possible rotation hypothesis:

$$f(R) = C(I_1, \Lambda_R I_2) \tag{1}$$

I_1 and I_2 are the two spherical images under consideration, Λ_R is an operator which rotates an image, and C determines the similarity between two images. In this setting, the rotation R which identifies the global maximum of the function f is the desired rotation. Of course, we must explicitly define the terms we have introduced here: A spherical image will be considered a function on the unit sphere parameterized with the spherical coordinates $\theta \in [0, \pi]$ (angle of colatitude) and $\phi \in [0, 2\pi)$ (angle of longitude). For utility, we define $\eta(\theta, \phi)$ to represent the unit vector associated with the spherical point $(\theta, \phi) \in \mathbb{S}^2$. We will use the notation $I(\eta(\theta, \phi))$ and $I(\theta, \phi)$ interchangeably. The rotation R is given as an element of the proper group of rotations $SO(3)$, which we will parameterize with the ZYZ Euler angles α, β, and γ. That is to say, any rotation can be written as

$$R(\alpha, \beta, \gamma) = R_z(\gamma) R_y(\beta) R_z(\alpha), \qquad \alpha, \gamma \in [0, 2\pi), \beta \in [0, \pi]$$

where R_z and R_y represent respectively rotations about the Z and Y axes. The rotation of an image I is then simply a rotation of image points by the operator Λ_R: $\Lambda_R I(\eta) = I(R^T \eta)$.

The critical step in our formulation is how we define the similarity between two images. We choose to identify two images if they are highly correlated. Equation (1) can be written concretely as [1]

$$f(R) = \int_{\eta \in \mathbb{S}^2} I_1(\eta) I_2(R^T \eta) d\eta \tag{2}$$

[1] If we choose to use normalized correlation to measure image similarity, we can simply replace the images $I_1(\eta), I_2(\eta)$ with $I_1(\eta) - \overline{I_1}, I_2(\eta) - \overline{I_2}$, and divide by the normalizing term $\sqrt{\int (I_1(\eta) - \overline{I_1})^2 \int (I_2(\eta) - \overline{I_2})^2}$. Since we are considering the entire signals for correlation, we can omit the dividing term. Here $\overline{I}$ is the image mean.

Essentially $f(R)$ computes the relative likelihood of each rotation in $SO(3)$. In principle this approach does not deviate from standard techniques applied to problems of pattern matching [4]. In such problems the search is for a planar shift (translational and/or rotational) which aligns a template pattern with a query image, where the correct alignment is given as the location of highest correlation. Such a formulation exploits general convolution principles which replace convolution with multiplication in the Fourier spectrum. An analogous computational speedup can be realized if we recognize that analyzing the spectral information of images defined on planes or spheres is part of the same general framework of harmonic analysis on homogeneous spaces. As we will see in the following sections, the Fourier coefficients of our correlation function $f(R)$ in (2) can be obtained from the pointwise multiplication of the Fourier coefficients of our two spherical images I_1 and I_2.

2.1. The Spherical Fourier Transform

This treatment of spherical harmonics is based on [5, 1]. In traditional Fourier analysis, periodic functions on the line (or equivalently functions on the circle $\mathbb{S}^1$), are expanded in a basis spanned by the eigenfunctions of the Laplacian. Similarly, the eigenfunctions of the spherical Laplacian provide a basis for $f(\eta) \in \mathcal{L}^2(\mathbb{S}^2)$ (here $\mathcal{L}^2(\mathbb{S}^2)$ denotes square-integrability, meaning the set of functions f such that $\int |f(\eta)|^2 d\eta$ is finite). These eigenfunctions are the well known spherical harmonics ($Y_m^l : \mathbb{S}^2 \mapsto \mathbb{C}$), which form an eigenspace of harmonic homogeneous polynomials of dimension $2l+1$. Thus, the $2l+1$ spherical harmonics for each $l \geq 0$ form an orthonormal basis for any $f(\eta) \in \mathcal{L}^2(\mathbb{S}^2)$. The $(2l+1)$ spherical harmonics of degree l are given as

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

where P_m^l are the associated Legendre functions and the normalization factor is chosen to satisfy the orthogonality relation

$$\int_{\eta \in S^2} Y_m^l(\eta) Y_{m'}^{l'}(\eta) d\eta = \delta_{mm'} \delta_{ll'}, \tag{3}$$

where δ_{ab} is the Kronecker delta function. Any function $f(\eta) \in \mathcal{L}^2(\mathbb{S}^2)$ can be expanded in a basis of spherical harmonics:

$$f(\eta) = \sum_{l \in \mathbb{N}} \sum_{m=-l}^{l} \hat{f}_m^l Y_m^l(\eta) \;\; \text{where } \hat{f}_m^l = \int_{\eta \in S^2} f(\eta) \overline{Y_m^l}(\eta) d\eta \tag{4}$$

The $\hat{f}_m^l$ are the coefficients of the Spherical Fourier Transform (SFT). Henceforth, we will use $\hat{f}^l$ and Y^l to annotate vectors in $\mathbb{C}^{2l+1}$ containing all coefficients of degree l.

2.2. Fourier Transform on the Rotation Group and Correlation

As our likelihood function $f(R)$ in (2) is defined on $SO(3)$, and because we are seeking a convolution-like property describing the Fourier transform of $f(R)$, we need a Fourier transform on $SO(3)$ in addition to a spherical transform. Similar to the process in the previous section, we can develop a Fourier transform on the rotation group $SO(3)$. When considering functions $f \in \mathcal{L}^2(SO(3))$, the Fourier transform can be described as a change of basis from the group elements to the basis of irreducible matrix representations. The spherical harmonic functions Y_m^l form a complete, orthonormal set providing a basis for the representations of SO(3). Furthermore, Schur's First Lemma from fundamental representation theory shows that they also supply a basis for the irreducible representations of $SO(3)$:

$$\Lambda_R Y^l(\eta) = U^l(R) Y^l(\eta). \tag{5}$$

The matrix elements of U^l are given by [4]

$$U_{mn}^l(R(\alpha,\beta,\gamma)) = e^{-im\gamma} P_{mn}^l(\cos(\beta)) e^{-in\alpha} \qquad m,n = -l,\ldots,l. \tag{6}$$

The P_{mn}^l are generalized associated Legendre polynomials which can be calculated efficiently using recurrence relations. Such an Euler angle parameterization of the irreducible representations of $SO(3)$ leads to a useful expansion of functions $f \in \mathcal{L}^2(SO(3))$:

$$f(R) = \sum_{l\in\mathbb{N}} \sum_{m=-l}^{l} \sum_{p=-l}^{l} \hat{f}_{mp}^l U_{mp}^l(R) \quad \text{where } \hat{f}_{mp}^l = \int_{R\in SO(3)} f(R)\overline{U_{mp}^l(R)} dR \tag{7}$$

The $\hat{f}_{mp}^l$, with $m,p = -l,\ldots,l$ are the $(2l+1)\times(2l+1)$ coefficients of degree l of the $SO(3)$ Fourier transform (SOFT).

As we are interested in relating two images separated by a rotation in Fourier space, we must firmly understand the effect of 3D rotations in this space. Intuitively, we would expect a rotation to manifest itself as a modulation of the Fourier coefficients as is the case in traditional Fourier analysis. This is, in fact, the observed effect. As spherical functions are rotated by elements of the

rotation group $SO(3)$, the Fourier coefficients are "modulated" by the irreducible representations of $SO(3)$:

$$f(\eta) \mapsto f(R^T\eta) \Longleftrightarrow \hat{f}^l \mapsto U^l(R)^T \hat{f}^l \tag{8}$$

In effect, the U^l matrix representations of $SO(3)$ are the spectral analogue to 3D rotations. The unitarity of these representations ensures that the rotation of a function does not alter the distribution of spectral energy among degrees:

$$||U^l(R)\hat{f}^l|| = ||\hat{f}^l||, \forall R \in SO(3) \tag{9}$$

We may now return our attention to the problem of estimating the rotation between two spherical images. Examining (2) more closely, we have developed the necessary tools to treat both $I_1(\eta)$ and $I_2(R^T\eta)$ with their respective Spherical Fourier expansions. Recently, [13, 19, 7] have explored the computation of such a correlation in the spectral domain. In the following proposition, we have replaced image labels I_1 and I_2 from (2) with the more generic labels $g(\eta), h(\eta) \in \mathcal{L}^2(\mathbb{S}^2)$.

Proposition 1. *If $f(R) = \int g(\eta)h(R^T\eta)d\eta$, such that $g, h \in \mathcal{L}^2(\mathbb{S}^2),\ f(R) \in \mathcal{L}^2(SO(3))$, then $f(R) = \sum_l (\hat{g}^l)^T \overline{U^l(R)\hat{h}^l}$, and the $SO(3)$ Fourier transform of $f(R)$ is given by $\hat{f}^l_{mp} = \hat{g}^l_m \overline{\hat{h}^l_p}$.*

Proof. Expanding the integral $\int g(\eta)h(R^T\eta)d\eta$ with the Fourier expansions of both $g(\eta)$ and $h(R^T\eta)$, we see immediately

$$f(R) = \sum_l \sum_{m=-l}^{l} \sum_n \sum_{p=-n}^{n} \sum_{k=-n}^{n} \hat{g}^l_m \overline{\hat{h}^n_p U^n_{pk}(R)} \int_{\eta\in\mathbb{S}^2} \overline{Y^n_k(\eta)} Y^l_m(\eta) d\eta.$$

Given the orthogonality of the spherical harmonic functions (3), the only nonzero terms in the summation appear when $n = l$ and $k = m$, so

$$f(R) \ = \ \sum_l \sum_{m=-l}^{l} \sum_{p=-l}^{l} \hat{g}^l_m \overline{\hat{h}^l_p U^l_{pm}(R)}. \tag{10}$$

This completes the first proof. The second statement requires the $SO(3)$ Fourier transform of $f(R)$. A direct application of the SOFT (7) produces

$$\hat{f}^n_{qr} = \sum_l \sum_{m=-l}^{l} \sum_{p=-l}^{l} \hat{g}^l_m \overline{\hat{h}^l_p} \int_{R\in SO(3)} \overline{U^l_{pm}(R)} U^n_{qr}(R) dR$$

The orthogonality of the matrices $U^l(R)$ ($\int U^l_{mp}(R)\overline{U^n_{qr}(R)}dR = \delta_{ln}\delta_{mq}\delta_{pr}$) yields nonzero terms in the summation only when $l = n$, $m = q$, and $p = r$, resulting in the reduced form

$$\hat{f}^l_{mp} \;=\; \hat{g}^l_m\overline{\hat{h}^l_p} \tag{11}$$

and thus completing our proof. □

As we had initially desired, the correlation of two spherical functions reflects the similar properties of a generalized convolution: the $SO(3)$ Fourier coefficients of the correlation of two spherical functions can be obtained directly from the pointwise multiplication of the individual SFT coefficients. In vector form, the $(2l+1) \times (2l+1)$ matrix of SOFT coefficients $\hat{f}^l$ is equivalent to the outer product of the coefficient vectors $\hat{g}^l$ and $\hat{h}^l$. Given $\hat{f}^l$, the inverse SOFT retrieves the desired function $f(R)$ with $(2L+1)$ samples in each of the three Euler angles, leaving us with accuracy up to $\pm\left(\frac{180}{2L+1}\right)^\circ$ in α and γ and $\pm\left(\frac{90}{2L+1}\right)^\circ$ in β (Here L is the bandwidth of the spherical images, meaning the largest degree for which we retain SFT coefficients).

To compute the SFT of a discrete spherical image, we can use a fast $O(L^2log^2L)$ algorithm developed by Driscoll and Healy [5], where L is the bandwidth of the signal being transformed. A similar separation-of-variables approach exists for a fast $SO(3)$ Fourier transform in $O(L^3log^2L)$ [13].

To recap the developments in this section, we have described an efficient method to estimate the rotation between two spherical images. Recall that the correlation function $f(R)$ in (2) computes the relative likelihood of each 3D rotation, and the maximum of this grid indicates the rotation separating the two images. The values of the function $f(R)$ can be obtained by first taking (discrete) Spherical Fourier transforms of the two images under consideration, and then taking an (discrete) inverse $SO(3)$ Fourier transform of their product. A search through these function values will yield the global maximum. While a direct computation of $f(R)$ would have a complexity of $O(L^3N)$ (where N, the number of spherical image pixels, must be at least $4L^2$), our algorithm is dominated only by the inverse $SO(3)$ Fourier transform.

2.3. DFTs Versus SOFT

An alternative to obtaining the correlation function $f(R)$ in (2) via an inverse SOFT is to generate the function samples using only inverse discrete Fourier transforms. This approach has been de-

tailed in [19, 14] and we recap it here. In short, the unitarity of the representations U^l allow for the decomposition of any rotational modulation term:

$$R_2 R_1 = R \Longrightarrow U^l(R_2)U^l(R_1) = U^l(R) \tag{12}$$

Using the rotational decomposition $R(\alpha, \beta, \gamma) = R_2(\beta + \pi, \frac{\pi}{2}, \gamma + \frac{\pi}{2})R_1(\alpha + \frac{\pi}{2}, \frac{\pi}{2}, 0)$ along with (12), we can follow the approach of the previous section to arrive at

$$f(R) = \sum_l \sum_{m=-l}^{l} \sum_{p=-l}^{l} \sum_{k=-l}^{l} \hat{g}^l_m \overline{\hat{h}^l_p} P^l_{pk}(0) P^l_{km}(0) e^{-i(p\gamma' + k\beta' + m\alpha')} \tag{13}$$

where $\gamma' = \gamma + \frac{\pi}{2}, \beta' = \beta + \pi, \alpha' = \alpha + \frac{\pi}{2}$. The Euler angles of rotation all appear in exponentials due to the decomposition of representations (12). The 3D FFT of $f(R)$ is easily computed as

$$\hat{f}_{mpk} = \sum_l \hat{g}^l_m \overline{\hat{h}^l_p} P^l_{pk}(0) P^l_{km}(0).$$

This result shows we can compute the discrete Fourier transform of our likelihood grid $f(R)$ as a combination of spherical Fourier coefficients. The computational cost added by summing through the degrees l is offset because the function samples can be computed with inverse FFTs for all three Euler angles.

2.4. Rotation Estimation from Conservation Constraints

Although the correlation technique described in the previous sections generates a likelihood measure for all possible rotation hypotheses, it provides little flexibility for a refined solution. As the Fourier transform is a global integral transform, achieving a finer resolution in rotation space requires selecting a greater bandwidth L for the SOFT and SFT transforms. An alternative would be to estimate the parameters of rotation directly from the spherical harmonic coefficients. Recall that the shift theorem (8) constrains the relationship between the SFT coefficients of rotated images. Another way of stating this result is with the following equality. For the correct aligning rotation R,

$$\|\hat{g}^l - U^l(R)^T \hat{h}^l\| = 0, \forall l \in \mathbb{N} \tag{14}$$

If we examine this statement further, we would see that the zero-order coefficients are unaffected by rotations about the Z axis (i.e. if $g(\eta) = \Lambda_{R_z} h(\eta) \Longrightarrow \hat{g}^l_0 = \hat{h}^l_0$). For a full rotation ($g(\eta) =$

$\Lambda_{R(\alpha,\beta,\gamma)}h(\eta)$), the zero-order coefficients are not altered by the final rotation γ:

$$\hat{g}_0^l - \left[U^l(R(\alpha,\beta,0))^T\hat{h}^l\right]_0 \quad = \quad 0, \forall l \in \mathbb{N} \tag{15}$$

Since (15) is true for any 3D rotation $R(\alpha,\beta,\gamma)$, the Euler angles α and β can be estimated independently of γ by minimizing the residual error (this equation can be further decomposed by expanding the representations using (12)). Subsequently, after the SFT coefficients $\hat{h}^l$ have been de-rotated with the estimates for α and β, the remaining angle γ can be determined from the following constraint which holds for all rotations about the Z axis:

$$|\hat{g}_m^l - e^{-im\gamma}\hat{h}_m^l| \quad = \quad 0, \forall l \in \mathbb{N}, |m| \leq l \tag{16}$$

A natural question arises here. Minimizing the residual error between coefficients will provide a faster convergence versus searching throughout a global likelihood grid. Why can such a minimization only be used as a refinement? The answer comes from subtle distinctions in the assumptions made by the correlation and minimization approaches, which we will denote the *correlation* and *conservation hypotheses*, respectively. The premise of a minimization is that for the correct rotational alignment the SFT coefficients of two signals will be *equal*. This is much stronger than expecting them to be *highly correlated*. In the ideal case of complete spherical images, we may observe similar results. However, in practice we expect the spherical images to be *corrupted*: this can be caused by many factors, including a limited field-of-view, a dynamic scene, and sensor noise. To illustrate the potential differences between these approaches, we have compared the possible effectiveness of a residual minimization to that of a global likelihood evaluation in the following section.

3. Experiments

In the previous sections we presented two methods for rotation estimation. The first, based on the *correlation hypothesis*, generates a global likelihood grid for $SO(3)$. The second, based on the *conservation hypothesis*, minimizes the residual error between SFT coefficients. While both approaches can be expected to perform equally well in an ideal setting, this may not be the case when the camera's imaging surface does not cover the entire sphere. In practice, our system consists of a Canon Powershot G2 digital camera fastened to a parabolic mirror attachment from

RemoteReality^{TM}[20]. The mirror's field-of-view is 212° so the camera captures slightly more than a hemisphere of information. It is well known that a central catadioptric projection is equivalent to a spherical projection followed by a projection onto the plane [8], so we are able to map images from this system to a regular spherical grid. Figure (1) shows a sample catadioptric image obtained from a parabolic mirror and its corresponding projection onto the sphere. We will proceed

Figure 1: Left: a parabolic catadioptric image. Middle: the corresponding spherical image on a uniformly sampled spherical grid. Right: the spherical image as it would appear on the surface of the sphere

by comparing the validity of the *correlation* and *conservation* hypotheses in the presence of signal alterations caused by having only partial spherical images. Since the portion of the sphere imaged by an omnidirectional sensor is hemispherical, a rotation about the camera's Y axis (orthogonal to the optical axis) introduces new signal content. As the magnitude of this rotation increases, the image overlap lessens, and the premise that we are dealing with two images identical up to a 3D rotation is tested. We record the results from a global likelihood grid obtained with a bandwidth of $L = 128$. Here we perform the correlation three times by compensating for the rotation estimate after each iteration. Figure (2) shows the results after the first and third iteration for each test. Even as the rotation angle β becomes obscenely large causing significant signal alterations, the correlation estimate accurately determines the rotation angle. It is interesting to note that for all the tests the small estimation errors are underestimation errors for the angle β. We perform a similar test to observe the accuracy of this approach for a wide range of function bandwidths L. As the results in Figure (3) indicate, the correlation estimate is quite accurate even when we retain few coefficients from the spherical harmonic expansions, which is quite useful for time critical applications.

To test the stability of the *conservation hypotheses* we simulate small rotations of a catadioptric sensor. In lieu of performing the minimization, we searched in a large neighborhood of the ground-truth solution to verify that the rotation minimizing the residual error was indeed the ground-truth

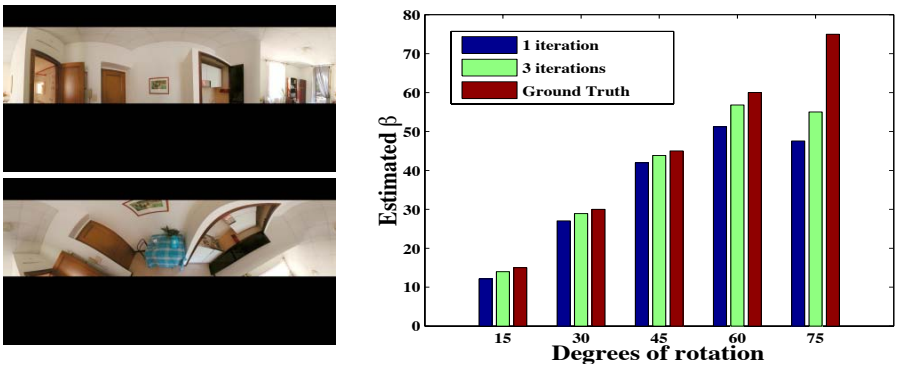

Figure 2: Left: A pair of simulated catadioptric images related by a rotation about the Y axis of $60°$. Right: Correlation results in estimating the Euler angle β under large rotations of a catadioptric sensor. Results for one and three iterations of the correlation search for five tests are shown. The ground-truth rotation about the Y axis ranges from $15°$ to $75°$ in increments of $15°$. Even up to $60°$ the correlation method provides a successful initial estimate.

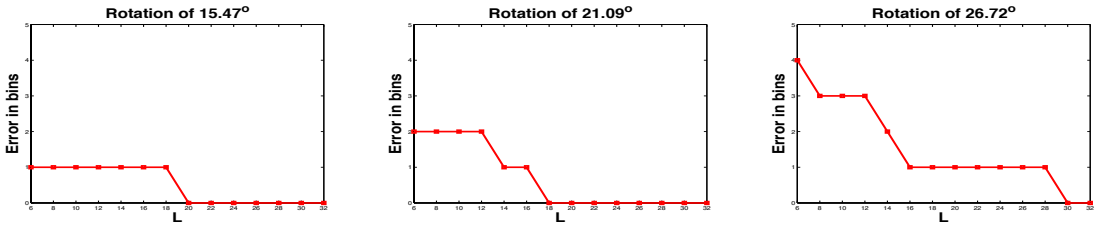

Figure 3: Errors in correlation estimates for function bandwidths L ranging from 6 to 32. Three artificial omni-image pairs were generated (with $15.5°$, $21.1°$, and $26.7°$ of rotation about the Y axis) in the same fashion as the test images used in Figure (2). Each plot shows the distance of the observed peak location from the ground truth in the estimated likelihood grid for a range of bandwidths. This error distance is measured in terms of bins (or samples) in the likelihood grid, where each bin represents $\frac{180°}{2L}$ (e.g. for $L = 6$, the bin centers are $15°$ apart).

rotation. To control this experiment we consider a spherical function with a known bandwidth. We can easily generate such a function from a linear combination of spherical harmonics:

$$g(\eta) = \Re\left[\sum_{l=0}^{L}\sum_{m=-l}^{l} kY_m^l(\eta)\right], \text{where } \Re[x] = \text{the real part of } x$$

Here k is a random number. We fix the bandwidth to $L = 15$, which assures that there are no unaccounted natural frequencies which may affect the computation. Since the minimization proposed in Section 2.4 is overconstrained, we can choose which SFT coefficients to use

($ERR = \sum_{l=L_1}^{L_2} \|\hat{g}^l - U^l(R)\hat{h}^l\|^2$). We performed this experiment for all $L_1 = 1 \ldots 15$ and $L_2 = L_1 \ldots 15$. A few representative samples from this set are shown in Figure (4). From the sam-

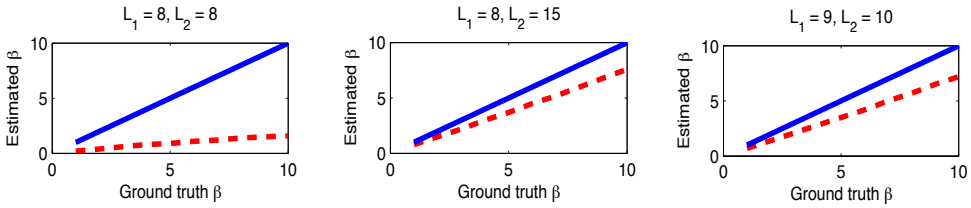

Figure 4: Representative plots depicting the rotations minimizing the residual error in different ranges of SFT coefficients. The solid line on the diagonal marks the position of the true rotation, while the dashed line marks the rotation minimizing the residual error. Due to signal corruption, β will be underestimated even for small rotations.

ple plots shown in the figure, it is clear that as the angle of rotation around the Y axis increases, the signal deterioration caused by a limited field of view may cause instabilities in any approach derived from the *conservation hypotheses*. These stability tests show that a nonlinear minimization of the residual error, because of its efficiency in computation, can be effective when the known rotation is small, but a search through a global likelihood grid will be capable even for extremely large rotations up to 60°.

3.1. Real rotations

The physical camera rotation where we can most closely monitor the ground truth is a rotation around the camera's optical axis. Here we have taken a sequence of 8 images where each consecutive pair is separated by a rotation of approximately 45° about the camera's optical (Z) axis. The rotation estimates obtained by generating a global likelihood grid are shown in Figure (5). From what was observed as the ground truth, our algorithm successfully estimated the rotation angle around the optical axis as well as correctly finding only a negligible rotation around the Y axis.

Figure (6) displays the results of our correlation approach on images obtained after arbitrary 3D rotations of a catadioptric sensor. In these images, the scene depth in many directions is quite small, and so there is inevitably a translational component to the motion as well. In the presence of a small translational motion and a hemispherical field of view, the global maximum of the

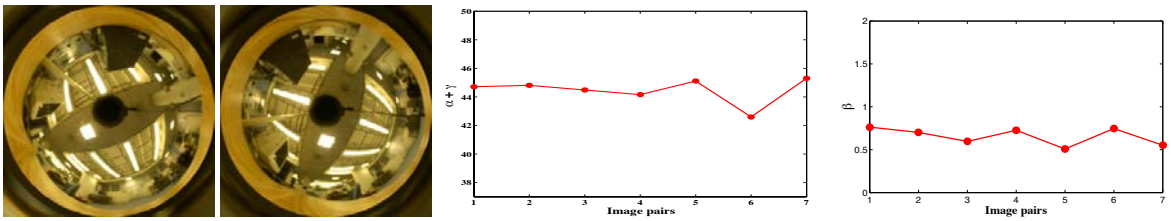

Figure 5: **A,B**: One of seven sequential pairs of images tested. The ground truth rotation was observed to lie at approximately 45° about the Z axis for each pair. **C**: The estimated rotation about the Z axis ($\alpha + \gamma$). **D**: The estimated rotation about the Y axis (β). Our algorithm correctly finds the rotation about the Z axis as well as finding only a negligible rotation about the Y axis.

correlation grid successfully recovers the dominant rotational motion.

4. Conclusion

We proposed a novel scheme for computing large rotations from spherical images without correspondences. The method has applications in omnidirectional images and navigation as well as to the alignment of 3D models. We also showed in [17] that it is quite resistant to small translations of a camera. The magnitude of the coefficient vector for a particular order can also be used as a rotation invariant to decide whether we have been at the same viewpoint or not. In our ongoing work [18], we are considering two calibrated spherical views without correspondences but satisfying the epipolar constraint. As opposed to a pure rotation, the transformation from one image to the other is depth dependent, thus allowing only the application of the epipolar constraint. We build a Hough space of rotations and baseline directions where each cell (motion hypothesis) receives a vote of all possible matches fulfilling the epipolar constraints. This vote can be written as a Radon integral and the challenge is how to compute this integral in a fashion similar to the methodology of this paper.

References

[1] G.B. Arfken and H.J. Weber. *Mathematical Methods for Physicists*. Academic Press, 1966.

[2] J.R. Bergen, P. Anandan, K.J. Hanna, and R. Hingorani. Hierarchical Model-Based Motion Estimation. In *Proc. Second European Conference on Computer Vision*, pages 237–252,

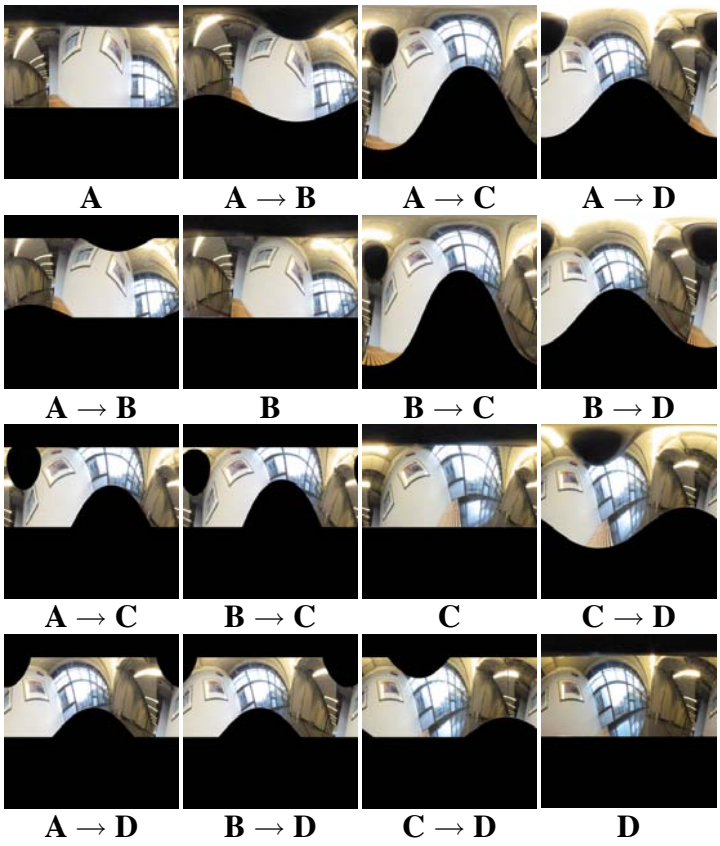

Figure 6: On the diagonal are four original catadioptric images (**A, B, C, D**). Because no ground-truth data is provided, we test our algorithm with visual alignment. To the right of each original image is its warped version showing alignment (e.g. the image marked **A**→**C** in the first row is image **A** rotated to align with image **C** after estimation). To the bottom of each original are the same warped images, but only showing the regions which are seen from both original catadioptric images. The correlation grid appears to produce very accurate results even with large rotations (e.g. the angle β was estimated at 46° between images B and C).

Santa Margerita, Italy, May 23-26, G. Sandini (Ed.), Lecture Notes in Computer Science 588, Springer-Verlag, Berlin et al., 1992.

[3] G. Burel and H. Henocq. Determination of the Orientation of 3D Objects Using Spherical Harmonics. *Graphical Models and Image Processing*, 57:400–408, 1995.

[4] G.S. Chirikjian and A.B. Kyatkin. *Engineering Applications of Noncommutative Harmonic Analysis: With Emphasis on Rotation and Motion Groups*. CRC Press, 2000.

[5] J.R. Driscoll and D.M. Healy. Computing Fourier Transforms and Convolutions on the 2-Sphere. *Advances in Applied Mathematics*, 15:202–250, 1994.

[6] C. Fermuller and J. Aloimonos. Direct Perception of Three-Dimensional Motion from Pat-

terns of Visual Motion. *Science*, 270:1973–1976, 1995.

[7] T. Funkhouser, M. Kazhdan, P. Shilane, P. Min, W. Kiefer, A. Tal, S. Rusinkiewicz, and D. Dobkin. Modeling by Example. In *Proceedings of ACM SIGGRAPH*, pages 652–663, 2004.

[8] C. Geyer and K. Daniilidis. Catadioptric Projective Geometry. *International Journal of Computer Vision*, 43:223–243, 2001.

[9] B. K. P. Horn. Extended Gaussian Images. *IEEE*, 72:1671–1686, 1984.

[10] S. B. Kang and K. Ikeuchi. The Complex EGI: A New Representation for 3-D Pose Determination. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(7):707–721, 1993.

[11] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors. In *Symposium on Geometry Processing*, June 2003.

[12] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Symmetry Descriptors and 3D Shape Matching. In *Symposium on Geometry Processing*, July 2004.

[13] P. J. Kostelec and D. N. Rockmore. FFTs on the Rotation Group. In *Working Paper Series, Santa Fe Institute*, 2003.

[14] J. A. Kovacs and W. Wriggers. Fast Rotational Matching. *Biological Crystallography*, 58:1282–1286, 2002.

[15] R. Lenz. Rotation-Invariant Operators. In *Proc. Int. Conf. on Pattern Recognition*, pages 1130–1132, Paris, France, Sept. 28-3, 1986.

[16] A. Leonardis and M. Jogan. Robust Localization Using Eigenspace of Spinning-Images. In *IEEE Workshop on Omnidirectional Vision, Hilton Head, SC, June 12*, pages 37–46, 2000.

[17] A. Makadia and K. Daniilidis. Direct 3D-Rotation Estimation from Spherical Images via a Generalized Shift Theorem. In *IEEE Conf. Computer Vision and Pattern Recognition*, Wisconsin, June 16-22, 2003.

[18] A. Makadia, C. Geyer, S. Sastry, and K. Daniilidis. Radon-Based Structure from Motion Without Correspondences. In *IEEE Conf. Computer Vision and Pattern Recognition*, San Diego, June 20-25, 2005.

[19] A. Makadia, L. Sorgi, and K. Daniilidis. Rotation Estimation from Spherical Images. In *Proc. Int. Conf. on Pattern Recognition*, Cambridge, UK, 2004.

[20] S. Nayar. Catadioptric Omnidirectional Camera. In *IEEE Conf. Computer Vision and Pattern*

Recognition, pages 482–488, Puerto Rico, June 17-19, 1997.

[21] S. Negahdaripour and B.K.P. Horn. Direct Passive Navigation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9:168–176, 1987.

[22] J. Sato and R. Cipolla. Extracting Group Transformations from Image Moments. *Computer Vision and Image Understanding*, 73:29–42, 1999.

[23] H. S. Sawhney and R. Kumar. True Multi–Image Alignment and its Application to Mosaicing and Lens Distortion Correction. In *IEEE Conf. Computer Vision and Pattern Recognition*, pages 450–456, Puerto Rico, June 17-19, 1997.

[24] J. Segman, J. Rubinstein, and Y. Zeevi. The Canonical Coordinates Method for Pattern Deformation: Theoretical and Computational Considerations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 14:1171–1183, 1992.

[25] H.Y. Shum and R. Szeliski. Systems and Experiment Paper: Construction of Panoramic Image Mosaics with Global and Local Alignment. *International Journal of Computer Vision*, 36:101–130, 2000.

[26] T.Pajdla and V.Hlavac. Zero Phase Representation of Panoramic Images for Image Based Localization. In *Int. Conf. on Computer Analysis of Images and Patterns*, pages 550–557, Ljubljana, Slovenia, Sept. 1-3, 1999.