

# A theory of phase-sensitive rotation invariance with spherical harmonic and moment-based representations

Ramakrishna Kakarala and Dansheng Mao  
School of Computer Engineering  
Nanyang Technological University  
{ramakrishna, DSMao}@ntu.edu.sg

## Abstract

*This paper describes how phase-sensitive rotation invariants for three-dimensional data may be obtained. A “bispectrum” is formulated for rotations, and its properties are derived for spherical harmonic coefficients as well as for moments. The bispectral invariants offer improved discrimination over previously published magnitude-only invariants. They are able to distinguish rotations from reflections, as well as rotations of an entire shape from component-wise rotations of elements of the shape. As experiments show, they provide robust performance for both surface and voxel data.*

## 1. Introduction

The classical approach to viewpoint-independent pattern recognition is by computing geometric invariants. Geometric invariants come in two types: spatial-domain, which are usually described as moments; and frequency-domain, which are usually described as magnitudes of Fourier or related transform coefficients. For example, to verify whether two patterns differ only by a three-dimensional (3-D) rotation, we may operate in the spatial domain and calculate the moment forms derived by Flusser *et al* [3]. Alternatively, we may operate in the frequency domain, and calculate the magnitudes of spherical harmonic coefficient vectors as shown by Kazhdan *et al* [12], or use the harmonic shape context of Frome *et al* [5]. In principle, the two types of geometric invariants are related through the generating equation of the relevant Fourier transform kernel, as established by Kanatani [11] for the sphere  $S^2$ .

Desirable characteristics of invariants include ease and stability of computation, and discriminative power to distinguish between similar-looking patterns that are not rotations of each other. Both moment forms and Fourier-domain magnitudes have their respective drawbacks in regard to those characteristics. A drawback to computing mo-

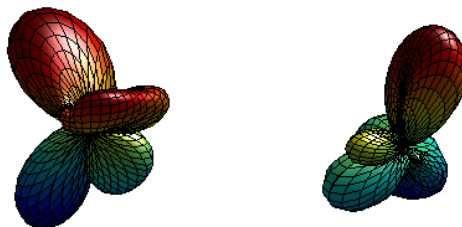


Figure 1. Two shapes that are not rotations of each other, and yet have identical spherical harmonic coefficient vector magnitudes. The shapes may be discriminated easily by using the bispectral invariants described in this paper.

ments, which are sums of the data multiplied by monomials such as, for example,  $x^2y^2z^2$ , is that they are not well-conditioned numerically. In general, the ratio of the maximum to minimum value of a monomial over a domain is  $O(B^p)$ , where  $B$  is the bound on the domain and  $p$  is the monomial’s order. A drawback to the Fourier-magnitude approach is that the magnitudes obviously lack the crucial phase information necessary to discriminate patterns. For example, the two functions shown in Figure 1 are not rotations of one another in 3-D, but have identical spherical harmonic magnitudes. Kazhdan *et al* [12] provide similar examples.

This paper describes a novel approach to generating phase-sensitive geometric invariants by using the bispectrum of group representations. Before going into groups or representations, the notions of phase-sensitivity and bispectrum may be grasped intuitively by considering the Fourier transform  $F$  of a function  $f$  on the real-line. It is obvious that at any frequency  $u$ , the magnitude  $|F(u)|$  is invariant to the translation of  $f$ . The magnitude is not a discriminating invariant, since it is actually phase that determines locations of edges and peaks. In contrast to a simple magnitude, consider for each pair of frequencies  $u, v$  the product known as the bispectrum  $B_f(u, v) := F(u)F(v)F(u+v)^*$ ; here  $*$  is complex-conjugate and  $:=$  means defined equal. The

bispectrum is not only invariant under the translation of  $f$  (as can be checked by replacing  $F(w)$  by  $F(w)e^{jwx}$ , where  $x$  is the amount by which  $f$  translates), but also contains nearly all other information in  $F$ . To see that, note that if we set  $v = 0$ , then  $B_f(u, 0) = F(0)|F(u)|^2$ , so the magnitude spectrum may be recovered from the bispectrum. Now, assuming that  $F$  never vanishes, we may recover the phase spectrum completely, up to a single linear phase factor corresponding to the translation, by recursively evaluating ratios such as  $B_f(u, 1)/(F(1)F(u))$ ; see [7] for algorithms and more details.

The bispectrum is well-known in statistical signal processing as a means to detect non-Gaussian frequency coupling. However, the bispectrum has not been used with the spherical harmonic representations employed in computer vision. This paper derives a convenient vector form and uses it to establish important properties of the bispectrum for 3-D rotation invariance. The principal advantage of the bispectral invariants over magnitudes is discriminative power due to phase-sensitivity. For example, as discussed in more detail below, the spherical harmonic bispectra of the two functions shown in Figure 1 are significantly different, enabling their discrimination. We show that, in general, the bispectrum is capable of distinguishing rotations from reflections, as well as rotations of an entire shape from component-wise rotations of elements of the shape.

The paper below begins with a review of the extensive literature on rotation invariants. Background material on spherical harmonics and group representations is provided subsequently. Next, the bispectrum of spherical harmonics is formulated and its significant properties are derived. We then illustrate the use of the bispectrum with four experiments on 3-D data, including data in surface form and in voxel form.

## 2. Review of literature on rotation-invariants

The use of 3-D rotation invariants is fundamental to many disciplines besides computer vision. For example, the physics literature discusses 3-D rotation invariants as quantities independent of the frame of reference, such as the addition formula for spherical harmonics [1, pg 213]. The rotation-invariance of such basic quantities as vector length or eigenvalues of a linear operator, and their related quantities (determinant, trace), is fundamental to linear algebra. Those concepts have appeared in the computer vision literature in various forms, as discussed below; see [4] for a thorough review.

Sadjadi & Hall [16] used the properties of the Hessian to obtain 3-D rotation invariants for quadratic equations in the coordinates  $x$ ,  $y$ , and  $z$ . Lo & Don [14] extended the development by using group representation theory, including the Clebsch-Gordan coefficients involved in reducing tensor products. The invariants that they propose are rela-

tively few, and are a special case of the invariants derived in this paper. Galvez & Canton [6] propose to rotate the shape into a canonical orientation where the principal axes (eigenvectors of the inertia matrix) are the Euclidean ones. However, if we have a complex object, such as MRI voxel data, rotating into a canonical orientation is computationally expensive, and the matching problem even if canonical orientation is provided is not simple.

In the literature on frequency-domain invariants, Reiser & Burkhardt [15] propose to use “bandwise” magnitudes of the spherical harmonic coefficients, to obtain improved discrimination over the vector-wise magnitudes of ref. [12]. The invariants we propose improve discrimination over bandwise magnitudes by allowing phase-sensitivity, where phase in this case means the corresponding element of the Wigner D-matrices described below.

## 3. Background material

The theory of 3-D rotations, group representations, and spherical harmonics is well-known. We review key concepts and results in this section; see ref. [8] for details. We use the  $z$ - $y$ - $z$  system of Euler angles, denoted  $\alpha$ ,  $\beta$ , and  $\gamma$ , to describe each 3-D rotation. The two angles,  $\alpha$  and  $\beta$ , also parameterize the sphere  $S^2$ , with  $0 \leq \alpha < 2\pi$  representing longitude (angle measured counter-clockwise from the positive  $x$ -axis in the  $x$ - $y$  plane),  $0 \leq \beta \leq \pi$  the colatitude (angle with respect to positive  $z$ -axis). In particular, we obtain the unit vector  $X$  defined as

$$X := [\cos(\alpha)\sin(\beta), \sin(\alpha)\sin(\beta), \cos(\beta)]^T. \quad (1)$$

Spherical harmonics are an orthogonal basis for functions on  $S^2$ , and, for each non-negative integer  $\ell$ , they have the functional form

$$Y_\ell^n(X) := Y_\ell^n(\beta, \alpha) = P_\ell^n(\cos\beta)e^{-jn\alpha}, \quad -\ell \leq n \leq \ell. \quad (2)$$

Here  $P_\ell^n$  are the *associated Legendre functions*, which are real-valued. Each square-integrable function  $f$  on the sphere may be expanded as follows:

$$f(X) := f(\beta, \alpha) = \sum_{\ell=0}^{\infty} \sum_{n=-\ell}^{\ell} F_\ell^n Y_\ell^n(\beta, \alpha) \quad (3)$$

We combine all coefficients for a given  $\ell$  into a  $1 \times (2\ell + 1)$  vector

$$\mathcal{F}_\ell := [F_\ell^{-\ell}, \dots, F_\ell^0, \dots, F_\ell^\ell]. \quad (4)$$

We refer to  $\mathcal{F}_\ell$  as the  $\ell$ -th Fourier coefficient of  $f$  below.

The rotation property of the spherical harmonic coefficients, which is discussed below, is crucial to the development of invariants. Every 3-D rotation is represented by  $3 \times 3$  orthogonal matrix  $R$  with determinant  $+1$ . If, for some  $R$ , we have  $g(X) = f(RX)$  for all  $X \in S^2$ , so that

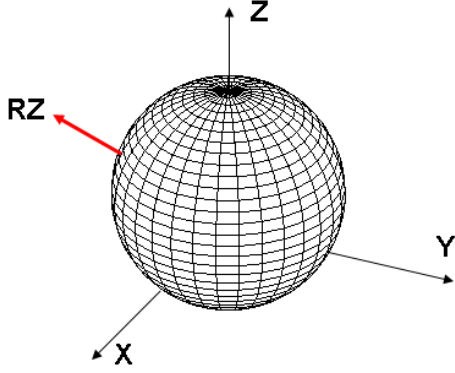


Figure 2. Illustration of the lifting process: for each 3-D rotation  $R$ , the value of  $\tilde{f}(R)$  is the value of  $f$  at  $RZ$ , the location of the  $Z$  axis after rotation by  $R$ .

$g$  is a rotated version of  $f$ , then the Fourier coefficients of  $g$  are related to those of  $f$  as follows:

$$\mathcal{G}_\ell = \mathcal{F}_\ell D_\ell(R). \quad (5)$$

Here  $D_\ell(R)$  is a  $(2\ell + 1)$ -dimensional unitary matrix, known as the Wigner  $D$ -matrix, with the homomorphic property  $D_\ell(RS) = D_\ell(R)D_\ell(S)$  for every pair of rotations  $R, S$ . The elements of the  $D$ -matrices are separable in the Euler angles:

$$D_\ell^{mn}(\alpha, \beta, \gamma) = e^{-jm\alpha} d_\ell^{mn}(\beta) e^{-jn\gamma}, \quad -\ell \leq m, n \leq \ell. \quad (6)$$

Here  $d_\ell^{mn}$  is the (little) Wigner  $d$ -function, which is real-valued. In particular, the spherical harmonics are, up to a scale factor, the elements of the middle column of the corresponding  $D$  matrix:

$$Y_\ell^m(\beta, \alpha) = a_\ell D_\ell^{m0}(\alpha, \beta, 0), \quad -\ell \leq m \leq \ell. \quad (7)$$

The properties of the spherical harmonic Fourier transform are often clearer when  $S^2$  is expressed as a homogeneous space for the group  $SO(3)$  of 3-D rotations. The group consists of all  $3 \times 3$  orthogonal matrices with determinant  $+1$ . Every function  $f$  whose domain is  $S^2$  may be “lifted” to a corresponding function  $\tilde{f}$  on  $SO(3)$  by the mapping  $\tilde{f}(R) := f(RZ)$ , where  $Z := [0, 0, 1]^T$ . Figure 2 illustrates the lifting operation.

Note that  $\tilde{f}$  is constant on rotations that fix  $Z$ , and, therefore,

$$\tilde{f}(\alpha, \beta, \gamma) = \tilde{f}(\alpha, \beta, 0), \quad \forall \gamma. \quad (8)$$

The elements  $\{D_\ell^{mn}\}_{\ell, m, n}$  of the Wigner  $D$ -matrices form an orthogonal basis for functions defined on  $SO(3)$ . The Fourier transform on  $SO(3)$  consists of matrix-valued coefficients, one at each “frequency”  $\ell$ , where  $\ell = 0, 1, \dots$ . The  $\ell$ -th Fourier coefficient is defined as

$$\mathcal{F}(\ell) := \int_{SO(3)} f(R) D_\ell(R)^\dagger dR. \quad (9)$$

Here the symbol  $\dagger$  denotes matrix hermitian and  $dR = \frac{1}{8\pi} \sin(\beta) d\alpha d\beta d\gamma$  is the normalized rotation-invariant measure on  $SO(3)$ . With those Fourier coefficients, the function  $f$  may be expanded as follows

$$f(R) = \sum_{\ell=0}^{\infty} (2\ell + 1) \text{trace}[\mathcal{F}(\ell) D_\ell(R)]. \quad (10)$$

A 3-D rotation of  $f$ , which equivalently translates the function on  $SO(3)$ , produces a corresponding “phase-shift” of the Fourier coefficients:

$$g(R) = f(SR) \Leftrightarrow \mathcal{G}(\ell) = \mathcal{F}(\ell) D_\ell(S). \quad (11)$$

If  $f = \tilde{f}$ , or in other words  $f$  is a function lifted from  $S^2$  as described above, then from eqns. (6) and (8) we obtain that only the middle row of the  $(2\ell + 1)$ -dimensional Fourier coefficient matrix  $\tilde{\mathcal{F}}(\ell)$  is non-zero.

$$m \neq 0 \Rightarrow \tilde{\mathcal{F}}(\ell)^{mn} = 0. \quad (12)$$

Moreover, from (7), we obtain that for lifted functions, the middle rows of the  $SO(3)$  Fourier coefficients are, up to a scale factor, the Fourier coefficients of the spherical harmonic expansion:

$$\tilde{\mathcal{F}}(\ell)^{0n} = a_\ell \mathcal{F}_\ell^n. \quad (13)$$

The interested reader may verify that with (13) and (7), the expansion (10) reduces to (3).

### 3.1. The bispectrum

The bispectrum is a well-known concept in the signal processing literature, but is not familiar in computer vision. The bispectrum, and its time-domain version, the triple correlation, are best introduced by considering functions on the real line. The triple correlation of a function  $f$  is a generalization of the familiar autocorrelation to two independent shifts:

$$T_f(t_1, t_2) := \int_{\mathbb{R}} f(x)^* f(x + t_1) f(x + t_2) dx. \quad (14)$$

The triple correlation is translation invariant: if  $g(x) = f(x - x_0)$ , then  $T_g = T_f$ , as a consequence of the invariance of the integral. The Fourier transform of the triple correlation  $T_f$  yields the bispectrum formula mentioned in the introduction:

$$B_f(u, v) = F(u)F(v)F(u + v)^*. \quad (15)$$

We now formulate the triple correlation and derive the bispectrum for the rotation group  $SO(3)$ . Some matrix terminology is helpful. The direct sum of two matrices  $A$  and  $B$  is denoted  $A \oplus B$ , and is the block-diagonal matrix with

$A$  in the upper-left corner and  $B$  in the lower-right corner. Similarly, the notation  $\bigoplus_{i=1}^n A_i$  means the block-diagonal matrix formed with  $A_1$  in the upper left corner, and continuing with  $A_2, \dots, A_n$ , occurring progressively lower and to the right. The tensor (or Kronecker) product of the  $n \times n$  matrix  $A$  and the  $m \times m$  matrix  $B$  is denoted  $A \otimes B$ , and is the structured  $nm \times nm$  matrix whose blocks are  $a_{ij}B$ , for  $1 \leq i, j \leq n$ . Note that  $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$ .

The triple correlation may be formulated on the rotation group  $SO(3)$  as follows:

$$T_f(S_1, S_2) := \int_{SO(3)} f(R)^* f(RS_1) f(RS_2) dR. \quad (16)$$

The triple correlation is therefore a function on  $SO(3) \times SO(3)$ . It is rotation invariant: if for some  $S$  we have  $g(R) = f(SR)$  for all  $R$ , then  $T_g = T_f$  as a consequence of the rotation invariance of the measure  $dR$ . To see that, note that for each  $\psi$  we have

$$\int_{SO(3)} \psi(SR) dR = \int_{SO(3)} \psi(R) dR. \quad (17)$$

The Fourier transform of the triple correlation, denoted  $B_f(k, \ell)$ , is expressed in terms of the tensor-product of the Wigner D-matrices:

$$\int \int T_f(S_1, S_2) [D_k(S_1) \otimes D_\ell(S_2)]^\dagger dS_1 dS_2. \quad (18)$$

The tensor product  $D_k \otimes D_\ell$  is unitarily equivalent to a direct sum when  $S_1 = S_2 = S$ , i.e.,

$$D_k(S) \otimes D_\ell(S) = C_{k\ell} \left[ \bigoplus_{i=|k-\ell|}^{k+\ell} D_i(S) \right] C_{k\ell}^\dagger. \quad (19)$$

The unitary matrix  $C_{k\ell}$  is called the Clebsch-Gordan (CG) matrix; see [9, pg 135]. Applying (19), we obtain a convenient formula for the bispectrum:

$$B_f(k, \ell) = [\mathcal{F}(k) \otimes \mathcal{F}(\ell)] C_{k\ell} \left[ \bigoplus_{i=|k-\ell|}^{k+\ell} \mathcal{F}(i)^\dagger \right] C_{k\ell}^\dagger. \quad (20)$$

The above result is also discussed in Kakarala *et al* [10] and in Kondor & Borgwardt [13]. For example  $B_f(1, 1)$  is the  $9 \times 9$  matrix

$$[\mathcal{F}(1) \otimes \mathcal{F}(1)] C_{11} [\mathcal{F}(0) \oplus \mathcal{F}(1) \oplus \mathcal{F}(2)] C_{11}^\dagger. \quad (21)$$

The CG coefficients may be calculated using well-known recursive formulas [1, pp 134-137], with a phase convention that results in  $C_{k\ell}$  being a real-valued, orthogonal matrix for each  $k, \ell$ . Computation of the entries in a CG matrix may be accomplished by using published routines in Matlab<sup>1</sup>, Mathematica, or MATPACK.

<sup>1</sup>In the experiments for this paper, we used the Matlab routines from <http://geoweb.princeton.edu/people/simons/software.html>.

## 4. Deriving rotation invariants from the bispectrum

In this section, we establish some important properties of the bispectrum. First, in order to simplify calculations for spherical harmonics, we define a vector bispectrum for functions defined on  $S^2$ . Then, we show how that bispectrum may be extended to various forms of 3-D data by using projections, as well as by using moments.

Let  $f$  be a function defined on the sphere  $S^2$ , with corresponding spherical harmonic Fourier coefficient vectors  $\mathcal{F}_\ell$  for  $\ell \geq 0$ . We construct the bispectrum for vectors following (20). The term  $\mathcal{F}_k \otimes \mathcal{F}_\ell$  is a  $1 \times (2k+1)(2\ell+1)$  vector. We construct a compatible vector out of each  $\mathcal{F}_i$ , for  $|k-\ell| \leq i \leq k+\ell$ , by “embedding” it in a  $1 \times (2k+1)(2\ell+1)$  vector, padded with zeros. Define

$$\widehat{\mathcal{F}}_i := [0, \dots, 0, \mathcal{F}_i, 0, \dots, 0]. \quad (22)$$

Here, the number of zeros preceding  $\mathcal{F}_i$  is  $n_p$ , and the number succeeding  $\mathcal{F}_i$  is  $n_s$ , where

$$n_p := \sum_{q=|k-\ell|}^{i-1} (2q+1); \quad n_s := \sum_{q=i+1}^{k+\ell} (2q+1). \quad (23)$$

Note that the length of  $\widehat{\mathcal{F}}_i$  is  $n_p + 2 * i + 1 + n_s = (2k+1)(2\ell+1)$ . Now define the vector bispectrum

$$b_f^{k\ell}(i) := [\mathcal{F}_k \otimes \mathcal{F}_\ell] C_{k\ell} \widehat{\mathcal{F}}_i^\dagger; \quad |k-\ell| \leq i \leq k+\ell. \quad (24)$$

where  $C_{k\ell}$  is the Clebsch-Gordan coefficient matrix in (19). Note that the vector bispectrum is a scalar for each  $i$ , unlike (20), which is a matrix. Henceforth, we use  $b_f^{k\ell}$  to mean the vector whose entries are (24), and  $b_f$  to denote the collection of bispectral vectors for all  $k, \ell$ . The construction of (24), in particular the vectors  $\widehat{\mathcal{F}}_i$ , is also obtained in a different manner: if we had extended  $f$  from  $S^2$  to  $\tilde{f}$  on  $SO(3)$ , and used the bispectrum (20), we would obtain the same set of scalars  $\{b_f\}$  as a consequence of (13). The properties of  $b_f$  are now established.

**Theorem 4.1.** *Let  $f$  be a function on  $S^2$ . Then, we have the following.*

(a) *If  $f$  is lifted to  $\tilde{f}$  on  $SO(3)$ , then the non-zero entries of the product  $B_{\tilde{f}}(k, \ell) C_{k\ell}$  are precisely the entries in  $b_f^{k\ell}$ , up to scalar constants which are independent of  $f$ .*

(b) *The vector bispectrum is invariant to the rotation of  $f$ :*

$$g(X) = f(RX) \Rightarrow b_g = b_f. \quad (25)$$

(c)  *$b_f^{k\ell}$  is essentially the same as  $b_f^{\ell k}$ : for each  $k, \ell$ , and  $i$ , there exists a constant  $\zeta$ , independent of  $f$ , such that*

$$b_f^{k\ell}(i) = \zeta b_f^{\ell k}(i). \quad (26)$$

*Hence, we need only evaluate  $b_f^{k\ell}$  for  $k \geq \ell$ .*

(d) The magnitude invariants are special cases:

$$b_f^{k0} = \mathcal{F}_0 \|\mathcal{F}_k\|^2. \quad (27)$$

(e) Reflection of a real-valued  $f$  results in complex-conjugation of  $b_f$ : if  $O$  is any orthogonal matrix with determinant  $-1$ , then

$$g(X) = f(OX) \Rightarrow b_g = b_f^*. \quad (28)$$

*Proof.* Note that  $B_f(k, \ell)C_{k\ell}$  is, from (20), the matrix

$$[\mathcal{F}(k) \otimes \mathcal{F}(\ell)] C_{k\ell} \left[ \bigoplus_{i=|k-\ell|}^{k+\ell} \mathcal{F}(i)^\dagger \right]. \quad (29)$$

By (13), there is exactly one non-zero row in  $\mathcal{F}(k) \otimes \mathcal{F}(\ell)$ , and, furthermore, also by (13), the non-zero columns in the direct sum are exactly  $a_i^* \mathcal{F}_i^\dagger$ . Consequently, part (a) follows, and the scale factors independent of  $f$  are  $a_k a_\ell a_i^*$ . Now, to prove (25), note that by (5) we have

$$\widehat{\mathcal{G}}_i = \widehat{\mathcal{F}}_i \left[ \bigoplus_{i=|k-\ell|}^{k+\ell} D_i(R) \right]. \quad (30)$$

Therefore,  $b_g^{k\ell}(i)$  is the scalar

$$\mathcal{F}_k \otimes \mathcal{F}_\ell \left\{ [D_k(R) \otimes D_\ell(R)] C_{k\ell} \bigoplus_i D_i(R)^\dagger \right\} \widehat{\mathcal{F}}_i^\dagger. \quad (31)$$

By (19), the term in brackets  $\{\cdot\}$  is  $C_{k\ell}$ , showing that  $b_g^{k\ell}(i) = b_f^{k\ell}(i)$ . To prove (26), note that the triple correlation of  $\tilde{f}$  (the lifted version of  $f$  on  $SO(3)$ ) is symmetric:  $T_{\tilde{f}}(S_1, S_2) = T_{\tilde{f}}(S_2, S_1)$ . Moreover, the representations  $D_k(S_1) \otimes D_\ell(S_2)$  and  $D_\ell(S_2) \otimes D_k(S_1)$  have the same trace and therefore are unitarily equivalent [8, pg 114], i.e., there exists a unitary matrix  $U$  such that

$$UD_k(S_1) \otimes D_\ell(S_2)U^\dagger = D_\ell(S_2) \otimes D_k(S_1). \quad (32)$$

Therefore,  $B_{\tilde{f}}(k, \ell) = UB_{\tilde{f}}(\ell, k)U^\dagger$ , and the result follows from part (a). To prove (27), note that  $\mathcal{F}_0$  is a scalar, equal to the mean value of  $f$  on  $S^2$ , and hence  $\mathcal{F}_k \otimes \mathcal{F}_0 = \mathcal{F}_0 \mathcal{F}_k$ ,  $C_{k0} = I$ , and  $\widehat{\mathcal{F}}_k = \mathcal{F}_k$ . To prove (28), note that if  $O_y := \text{diag}[1, -1, 1]$  is the  $3 \times 3$  diagonal matrix reflecting  $y$  to  $-y$ , then  $S := O_y O$  has determinant  $+1$  and therefore belongs to  $SO(3)$ . Hence, if  $h(X) := f(SX)$ , then  $b_h = b_f$  by (25). Now, note that if  $g(X) := h(O_y X) = f(OX)$ , then  $g(\beta, \alpha) = h(\beta, -\alpha)$  from (1), and hence by (3) we obtain that  $\mathcal{G}_\ell = \mathcal{H}_\ell^*$  as  $g, h$  are real-valued. Therefore (28) follows from (24) as  $C_{k\ell}$  is real-valued. Note also that (28) holds also for complex-valued functions if  $g(X) := f(OX)^*$ .  $\square$

Theorem 4.1(d) shows that the bispectral invariants  $b_f^{k\ell}(i)$  include, as a special case when  $\ell = 0$ , the magnitude-only invariants used in previous papers [12]. Moreover, for the case when  $\ell \neq 0$ , the bispectral invariants contain information about the phase spectrum of  $f$ , which is sufficiently discriminating to distinguish rotations from reflections by (25) and (28).

The bispectrum rotation invariants may also be extended to functions natively defined on  $SO(3)$ . The resulting bispectrum is matrix-valued, as shown on the right side of (20). Moreover, the properties established in the Theorem, in particular parts (b)-(e), extend with appropriate modifications to  $SO(3)$ . For example,  $B_f(k, 0) = \mathcal{F}(0)\mathcal{F}(\ell)\mathcal{F}(\ell)^\dagger$ , which is essentially a magnitude-only spectrum. The diagonal elements of  $B_f(k, 0)$  are the row-wise norms described in ref.[15]. Hence, for  $SO(3)$ , the bispectral invariants not only include the previously-published magnitude-based invariants as a special case, but more importantly, provide improved discrimination due to their phase sensitivity.

#### 4.1. Connection with moments

We now connect the bispectral invariants to the well-known theory of moments, which are a basic tool in pattern recognition. A moment of order  $\ell$  is the value of the integral

$$M_\ell^{m,n} := \int_{\mathbb{R}^3} x^m y^n z^{\ell-m-n} f(x, y, z) dx dy dz. \quad (33)$$

There are  $N_m(\ell) := (\ell + 1)(\ell + 2)/2$  moments of order  $\ell$ . Central moments are those taken after the function has been translated such that the 1-st order moment is zero; such a translation is  $x \mapsto x - M_1^{1,0}/M_0$ ,  $y \mapsto y - M_1^{0,1}/M_0$ , and  $z \mapsto z - M_1^{0,0}/M_0$ .

There exists a  $N_m(\ell) \times (2\ell + 1)$  matrix  $\mathcal{A}_\ell$  which transforms the vector of monomials of order  $\ell$  to the vector of spherical harmonics of the same order. The entries of  $\mathcal{A}_\ell$  may be found from the generating equation:

$$\frac{\sqrt{2\ell + 1}}{2^\ell t^\ell \ell!} (x - jy + 2zt - (x + jy)t^2)^2 = \sum_{n=-\ell}^{\ell} \frac{Y_\ell^n(\alpha, \beta) t^n}{\sqrt{(n+m)!(n-m)!}}$$

Here  $x, y$ , and  $z$  are the elements of the unit vector  $X$  in (1). For details behind those facts, see Kanatani [11, eq. 6.6], and for a proof, Courant-Hilbert [2, pp 540-541].

For example, let  $\ell = 2$ . Then, the vector of monomials is

$$\mathcal{M}_2 := [x^2, y^2, z^2, xy, yz, xz]. \quad (34)$$

The transforming matrix  $\mathcal{A}_2$  may be derived using the symbolic toolbox in Matlab, or equivalent package<sup>2</sup>. The values

<sup>2</sup>We used Matlab to calculate  $\mathcal{A}_\ell$  up to order 10 for the experiments described below.

shown below are for the first 3 rows of  $\mathcal{A}_2^\top$  (note the transpose). The values match the coefficients in [11, eq. 6.8].

$$\frac{\sqrt{30}}{4} \begin{bmatrix} 1 & -1 & 0 & 2i & 0 & 0 \\ 0 & 0 & 0 & 0 & -2i & 2 \\ -\frac{2}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{4}{\sqrt{6}} & 0 & 0 & 0 \end{bmatrix} \quad (35)$$

The remaining two rows are obtained by symmetry: the fourth row is the negative conjugate of the second, and the fifth row is the conjugate of the first. With this matrix, the row vector  $\mathcal{M}_2\mathcal{A}_2$  contains the spherical harmonics  $[Y_2^{-2}, Y_2^{-1}, \dots, Y_2^2]$ .

Therefore, the bispectral invariants that we derived for spherical harmonics may be applied to moments as well. We obtain the invariants

$$bm_f^{k\ell}(i) := [(\mathcal{M}_k\mathcal{A}_k) \otimes (\mathcal{M}_\ell\mathcal{A}_\ell)] C_{k\ell}(\widehat{\mathcal{M}_i\mathcal{A}_i})^\dagger. \quad (36)$$

A special case of this formula is when  $\ell = k$ , and  $i = 0$ . Detailed expressions for those particular invariants are in two other papers [3, eq 24][14, pg 1056].

## 5. Application to 3-D data

Data in 3-D in computer vision may come in several forms, including the following: in the form of  $(x, y, z)$  coordinates of vertices of polygonal shapes such as the PSB database [17]; or in the form of voxel data such as in MRI scans, which may be represented as a function  $f$  on a discrete grid in  $\mathbb{R}^3$ . We show in this section how to compute invariants for each of those forms of data.

Suppose we have a set of vertices, denoted  $\{(x_i, y_i, z_i)\}_{i=1}^N$ . Then we obtain the corresponding Fourier coefficients  $\mathcal{F}_\ell$  as follows. First, we convert each point  $(x_i, y_i, z_i)$  to spherical polar coordinates  $(r_i, \alpha_i, \beta_i)$ . Next, the Fourier coefficients are evaluated as follows:

$$F_\ell^m := \sum_{i=1}^N r_i Y_{\ell,m}(\beta_i, \alpha_i)^*. \quad (37)$$

We are essentially treating each  $r_i$  as the value of a “function”  $f$  at spherical angular coordinates  $(\beta_i, \alpha_i)$ . It is clear that when the set of vertices is rotated in 3-D, the coefficient vectors  $\mathcal{F}_\ell$  in (4) transform according to (5), and therefore the quantities  $b_f^{k\ell}(i)$  are invariant to 3-D rotation. Note that we may obtain combined translation, rotation, and uniform scale change invariants by translating the points by their centroid prior to calculating (37), and normalizing by  $(M_0^{00})^{\ell/3+1}$ ; see [14, eq. (31)].

Now suppose we have a second form of 3-D data, voxel data. For example, a MRI scan may be represented as  $f(x, y, z)$  for voxels whose coordinates are such that  $|x| \leq N_x$ ,  $|y| \leq N_y$ ,  $|z| \leq N_z$ . We may obtain the central moments of the scan from (33), and convert these to bispectral

invariants by (36). Alternatively, we may proceed solely in the frequency domain, and compute Fourier coefficients by converting each vector  $[x, y, z, f(x, y, z)]$  to polar coordinates, i.e.,  $[r, \alpha, \beta, f(r, \alpha, \beta)]$ , and using a version of (37) defined as follows:

$$F_\ell^m := \sum_{r,\alpha,\beta} r f(r, \alpha, \beta) Y_{\ell,k}(\beta, \alpha)^*. \quad (38)$$

Once again, we may compute the quantities  $b_f^{k\ell}(i)$ , and they are rotation invariants. In proceeding this way, we are essentially projecting the MRI data onto  $S^2$ , and computing the invariants of the projected data.

## 6. Experiments

In this section, we describe five experiments intended to establish the practical computation and reliability of the bispectral rotation invariants.

Our first experiment is to verify the ability of the bispectral invariants to discriminate rotation of the entire shape, from Fourier component-wise rotation of the elements of the shape. This experiment is motivated by one of the known failure modes of the magnitude invariants [12, Fig. 2]. We used shapes similar to those shown in Figure 1, each of which is generated by spherical harmonics using eq. (3) but having only 3 nonzero coefficient vectors:  $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ . Next, we generate a component-wise rotation by independently transforming each of the Fourier vectors as follows:

$$\mathcal{F}_1 \rightarrow \mathcal{F}_1 D_1(R_1); \mathcal{F}_2 \rightarrow \mathcal{F}_2 D_2(R_2); \mathcal{F}_3 \rightarrow \mathcal{F}_3 D_3(R_3). \quad (39)$$

When the rotations  $R_1, R_2$ , and  $R_3$  are different, the result is not a rotation of the entire shape. Magnitude-based invariants cannot detect component-wise rotation, but bispectrum invariants can. To verify that, we computed the  $1 \times 3$  vector  $b_f^{2,1}$  before and after (39). The normalized percent difference is computed as follows

$$E := \frac{\|b_f - b_g\|}{\|b_f\| + \|b_g\|} \times 100. \quad (40)$$

Here  $f$  is the original function, and  $g$  is the function after component-wise rotation in (39). Obviously, the larger  $E$  is, the more easily  $f$  and  $g$  are discriminated. Over 100 trials with randomly-generated rotations, the mean value of  $E$  was 72%, indicating very clear discrimination of component-wise rotation; in contrast, when a true rotation  $R_1 = R_2 = R_3 = R$  was tested with randomly chosen  $R$ , the values of  $E$  were essentially zero to machine precision.

Our second experiment was designed to verify the ability of the bispectrum to distinguish between rotation and reflection. We used the set of  $(x, y, z)$  vertices from a 3-D scan of a left hand<sup>3</sup>, shown in Figure 3. The reflection of the hand

<sup>3</sup>The hand shape appears courtesy of INRIA, and is available at the repository `shapes.aimatshape.net`.

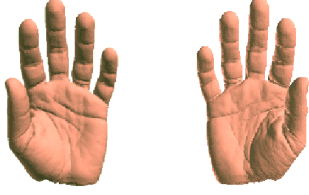


Figure 3. The vertices in the 3-D scan of the right hand were obtained by reflecting those of the left across the  $y$ - $z$  plane. While magnitude-only invariants can't tell the difference between the right and left hands, bispectral invariants can: the bispectral invariants of the right hand are complex conjugates of the left.

across the  $y$ - $z$  plane is also shown in the same figure. We computed the Fourier coefficients for  $\ell = 1, 2, \dots, 5$  of the hand vertices, where on each trial the vertices were reflected across a randomly-chosen plane in  $\mathbb{R}^3$ . We then compared the magnitude-based invariants  $\|\mathcal{F}_\ell\|$  of both the original and the reflected data sets. As expected, the magnitude-based invariants were essentially equal (to machine precision) for both the data and its reflection, despite there being no rotation to bring the two hands into alignment with one another. Next, the bispectral invariants  $b_f^{k\ell}(i)$  for  $k = 3$ ,  $\ell = 2$ , and  $i = 1, \dots, 5$  were computed. The normalized difference  $E$  for the two hands was approximately 40%, clearly indicating the ability of the bispectrum to distinguish reflections from rotations. However, when the bispectral values for the right hand were conjugated, they matched those of the left with essentially zero error  $E$ , confirming (28). The experiment was repeated with randomly chosen reflections, with no significant change in the results. Calculation of the invariants (using precomputed CG coefficients) required approximately 30 seconds in a Matlab program for a model containing over 53,000 vertices.

We repeated the experiment by selecting 39 shapes from the Princeton Shape Benchmark (PSB)[17] which were not bilaterally symmetric, such as human hands, walking figures, and potted plants. (Most shapes in PSB are bilaterally symmetric, because they represent balanced objects like aircraft or chairs). The 39 shapes chosen exhibit varying levels of asymmetry. Each shape was normalised in size to 1.0 mean distance from centroid, so that added noise may be expressed as a percentage. We reflected each shape across a randomly-chosen plane, and added Gaussian noise to the coordinates. We then calculated the distance (40) between the noise-free bispectral coefficients  $b_f$ , those of the noisy reflected shape, denoted  $\hat{b}_f$ , and their complex-conjugates  $\hat{b}_f^*$ . If the bispectral coefficients  $b_f$  are closer to  $\hat{b}_f$  than to  $\hat{b}_f^*$ , then we consider the shape to be confused with its reflection. With 5% noise, we found the confusion rate to be 2.6%; with 10% noise, the rate was 5.1%; and 15% noise resulted in confusion of 10.3%. Hence adding up to 10%

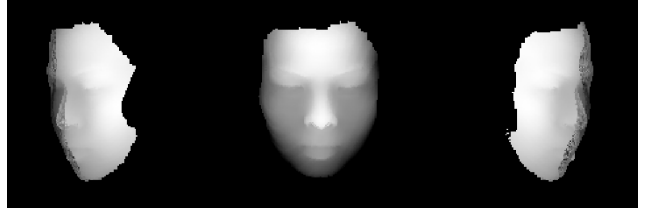


Figure 4. Three of the depth maps obtained from a real-time scanner. The subject moved his head from left to right, the extremes of which are shown. The bispectral invariants of each position were compared with the reference of the middle position shown.

noise does not incur significant error.

Our third experiment was designed to verify the robustness of the rotational invariants to rotation of a real-world object in 3-D. Depth maps were obtained of a rotating human face from a 3-D scanner<sup>4</sup>. The rotational motion is real in that it is due to the subject moving his face. Three images of the depth map are shown in Figure 4. The bispectral invariants  $b_f^{22}(i)$ , for  $0 \leq i \leq 4$ , were computed for 30 separate positions over the angular range of approximately  $-30$  to  $30$  degrees. The normalized difference  $E$  in invariants between each position and the approximately 0 degree position never exceeded 0.3%, with worst case occurring at the extremes of the rotation range. The bispectral invariants were computed both with the moment method (36), using central moments to obtain translation invariance, and with the spherical harmonic coefficient computed about the centroid using (37). There was no qualitative difference in the  $E$  values between the moment-based method and those obtained from (37), and no significant difference between using full resolution ( $480 \times 480$ ) and downsampled ( $48 \times 48$ ).

Our fourth experiment was designed to test the bispectral invariants on voxel data, and is similar to the experiment described in Flusser *et al* [3]. We used a MRI brain scan, originally  $256 \times 256 \times 60$  voxels. The central  $59 \times 59 \times 59$  voxels were extracted, and digitally rotated in all 3 Euler angles over 5 degree increments from 0 to 90 degrees. Figure 5 shows an example to illustrate the type of data used. The bispectral invariants  $b_f^{22}(i)$ ,  $0 \leq i \leq 4$  were computed by (38). The normalized error did not exceed 4%, with the worst error at the 45 degree rotation. The error is primarily due to aliasing as the rotated image is sampled on a discrete grid. Downsampling 3X in each dimension to speed up computation did not significantly change the results over full resolution.

Our fifth experiment tested how bispectral invariants perform when compared to magnitude invariants in shape recognition tasks. We used the 907 shapes in the PSB "testing" set [17], each randomly rotated with added noise over

<sup>4</sup>A Konica Minolta Vivid 910 non-contact digitizer was used to obtain the data.

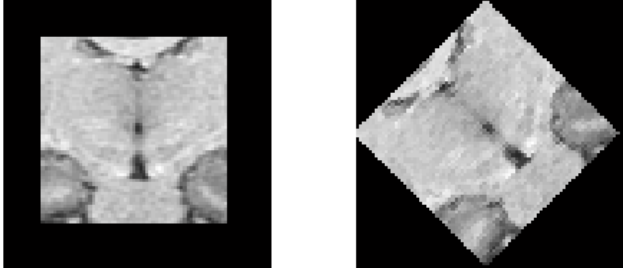


Figure 5. A slice of the MRI voxel data, taken perpendicular to the  $z$ -axis, is shown together with its 45-degree rotation. The bispectral invariants were evaluated as the MRI data was rotated in all 3 Euler angles.

trials, and computed the recognition rate (RR), which is how often the correct model is identified in noisy data. The shapes were normalised to mean distance of 1.0 from centroid as described above. We found that, in the presence of additive noise, RR improved significantly when bispectral invariants were used instead of magnitude invariants. Specifically, when 5% noise was added and a bandwidth of  $\ell \leq 7$  was used, we observed a RR of 96% for bispectral invariants, compared to 90% for magnitude invariants. Similarly, for 10% noise, the RR was 86% for bispectral invariants, compared to only 75% for magnitude invariants. Similar improvements were observed for other bandwidths, but are not reported here due to lack of space.

Both bispectral and magnitude invariants are computed from the same set of spherical harmonic coefficients. As the computation time of the invariants is negligible compared to that of the harmonic coefficients, we conclude that bispectral methods offer significant improvement in recognition tasks for relatively minor computation cost.

## 7. Conclusions

This paper proposes phase-sensitive 3-D rotation invariants by using the bispectrum of spherical harmonic coefficients. The invariants may be used for both surface and voxel data. Experiments show the ability of the invariants to distinguish rotations from reflections, as well as rotations of an entire shape from component-wise rotations of elements of the shape, while providing robust performance for matching rotated patterns.

## Acknowledgements

We thank the anonymous reviewers for their comments and the Institute of Media Innovation at NTU for funding this work with a Seed Grant.

## References

- [1] M. Chaichian and R. Hagedorn. *Symmetries in quantum mechanics: from angular momentum to supersymmetry*. Taylor& Francis, 1998.
- [2] R. Courant and D. Hilbert. *Methods of mathematical physics, pt I*. John Wiley, 1989.
- [3] J. Flusser, J. Boldys, and B. Zitová. Moment forms invariant to rotation and blur in arbitrary number of dimensions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):234–246, 2003.
- [4] J. Flusser, T. Suk, and B. Zitová. *Moments and moment invariants in pattern recognition*. Wiley, 2009.
- [5] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *ECCV (3)*, pages 224–237, 2004.
- [6] J. M. Galvez and M. Canton. Normalization and shape recognition of three-dimensional objects by 3d moments. *Pattern Recognition*, 26(5):667–681, 1993.
- [7] G. B. Giannakis. Signal reconstruction from multiple correlations: frequency- and time-domain approaches. *J. Opt. Soc. Am. A*, 6(5):682–697, 1989.
- [8] M. Hamermesh. *Group theory and its application to physical problems*. Dover, 1962.
- [9] E. A. Hewitt and K. A. Ross. *Abstract Harmonic Analysis, Vol. II*. Springer Verlag, 1970.
- [10] R. Kakarala, B. M. Bennett, G. J. Iverson, and M. D’Zmura. Bispectral techniques for spherical functions. In *Proceedings of ICASSP*, volume 4, pages 216–219, 1993.
- [11] K. Kanatani. Distribution of directional data and fabric tensors. *International Journal of Engineering Science*, 22(2):149–164, 1984.
- [12] M. M. Kazhdan, T. A. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3d shape descriptors. In *Symposium on Geometry Processing*, pages 156–165, 2003.
- [13] R. Kondor and K. Borgwardt. The skew spectrum of graphs. In A. McCallum and S. Roweis, editors, *Proceedings of the International Conference on Machine Learning*, pages 496–503. Omnipress, 2008.
- [14] C.-H. Lo and H.-S. Don. 3-d moment forms: Their construction and application to object identification and positioning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(10):1053–1064, 1989.
- [15] M. Reiser and H. Burkhardt. Using irreducible group representations for invariant 3d shape description. In *DAGM-Symposium*, pages 132–141, 2006.
- [16] F. A. Sadjadi and E. L. Hall. Three dimensional moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:127–136, 1980.
- [17] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The Princeton shape benchmark. In *In Shape Modeling International*, pages 167–178, 2004.