A novel set of rotationally and translationally invariant features for images based on the non-commutative bispectrum

Risi Kondor

risi@cs.columbia.edu

Computer Science Department, Columbia University,
1214 Amsterdam Ave., New York, NY10027, USA

Abstract

We propose a new set of rotationally and translationally invariant features for image or pattern recognition and classification. The new features are cubic polynomials in the pixel intensities and provide a richer representation of the original image than most existing systems of invariants. Our construction is based on the generalization of the concept of bispectrum to the three-dimensional rotation group SO(3), and a projection of the image onto the sphere.

1 Introduction

The representation of data instances in learning algorithms is subject to the conflicting demands of wanting to incorporate as much information as possible about real world objects, and not wanting to introduce spurious information with no physical meaning. Image recognition is perhaps the most striking example of this phenomenon: clearly, the position and orientation of an object inside a larger image is purely a matter of representation and not a property of the object itself.

There have been many attempts to construct rotation and translation invariant representations both in the vision community and in the machine learning world. A faithful representation of invariances is particularly important when pushing algorithms towards the limit of small training sets. When training data is abundant, it can drone out spurious degrees of freedom or average over them. However, in small datasets effective generalization is not possible without explicitly taking the invariances into account.

Various types of invariants are used in signal processing and computer vison, each with its own advantages and disadvantages (see, e.g., [9][7]). However, a common feature of most of these invariants is that they are lossy, in the sense that they do not uniquely specify the original data image. This becomes a particularly serious problem in discriminative learning, where the success of modern algorithms is to a large extent based on their ability to handle very high dimensional data, capturing as much information about data instances as possible. This is why in many cases (such as the character recognition problem to be addressed in the experimental section) it has often proven to be better to ignore the invariance altogether rather than risk losing valuable information as a side-effect of enforcing it.

Another potential problem with existing methods is their high computational cost. Approaches based on summing over members of the invariance group (ghost instances, etc.) and methods that require an expensive kernel evaluation for each pair of instances suffer specially badly from speed issues (e.g., [5]).

In this paper we propose a new class of invariant features for two dimensional images based on the algebra of generalized bispectra and a projection from the image plane onto the sphere. The new invariant features are strictly rotation and translation invariant (up to our bandwidth restriction and a small projection error), and close to complete, in the sense of uniquely specifying the original image up to a single rotation and translation. The bispectral invariants can be computed in a pre-processing step before any learning takes place in time $O(u^{5/2})$, where u is the size of the original image in pixels. The individual invariants are third order polynomials in the pixel intensities, and hence are relatively well behaved. We envisage the invariants to be used as inputs to an existing machine learning algorithm, for example as features to build kernels from. Our experiments show that using the bispectral invariants makes an immediate impact on a standard optical character recognition task when the training and testing intances are allowed to randomly translate and rotate.

While the bispectrum is well known in some areas

of vision and signal processing, most practicioners are only familiar with its classical "Euclidean" version [2]. For our purposes this is not sufficient because rotations and translations together form a non-commutative group. In particular, previous work on using the bispectrum for translation and rotation invariance considered these two types of transformations separately, first eliminating the unknown translation and then the rotation from the image [8]. While this is possible for image reconstruction, as regards generating invariant features it would amount to no more than transforming the image to a canonical position and orientation, which is obviosuly sensitive to variations in the image, since small changes can lead to vastly different optimal alignments with the canonical orientation.

While there is a well-developed and beautiful abstract theory of bispectra on general compact groups developed chiefly by Ramakrishna Kakarala [4] [3] [1], not many connections of the non-commutative case to real world problems have been explored. To the best of our knowledge, bispectra over non-commutative groups have never been used in the context of simultaneously enforcing rotational and translational symmetries of two-dimensional images. The crucial new device connecting rotations and translations of the plane to the action of a compact non-commutative group is the projection onto the sphere proposed in this paper.

The first half of this paper sets the scene by giving a rather abstract and general introduction to the theory of bispectra on groups. The second half of the paper contains our actual construction and the details of implementing it on a computer. The reader who is not interested in the wider context of bispectral invariants might find it convenient to skip directly to section 3.

2 Bispectral Invariants

The discrete **Fourier transform** of a complex-valued function $f\colon \{0, 1, 2, \ldots, n-1\} \to \mathbb{C}$ is defined

$$\widehat{f}(k) = \sum_{x=0}^{n-1} e^{-i2\pi xk/n} \, f(x), \qquad (1)$$

where k extends over $0, 1, 2, \ldots, n-1$ and each component $\widehat{f}(k)$ is the coefficient of the contribution to f at frequency k. A natural quantity of interest in signal processing is then the **power spectrum**

$$q(k) = \widehat{f}^*(k) \cdot \widehat{f}(k), \qquad (2)$$

where * denotes complex conjugation. The power spectrum quantifies how much energy the signal has in each frequency band. Intuitively it is clear that the power spectrum should be invariant to translations of

the signal. This is also borne out by the fact that by the convolution theorem the power spectrum is the Fourier transform of the **autocorrelation function**

$$\mathrm{corr}(x) = \sum_{y=0}^{n-1} f(y + x) \, f^*(y), \qquad (3)$$

(Wiener-Khinchin theorem) which is manifestly shift-invariant. Here and in the following addition and subtraction of indices and frequencies in $\{0, 1, 2, \ldots, n-1\}$ is always to be understood modulo n.

More formally, we define the **translate** of f by z as $f^z(x) = f(x - z)$. Plugging into (1),

$$\widehat{f^z}(k) = \sum_{x=0}^{n-1} e^{-i2\pi xk/n} \, f(x - z) =$$
$$\sum_{x=0}^{n-1} e^{-i2\pi(x+z)k/n} \, f(x) = e^{-i2\pi zk/n} \widehat{f}(k),$$
$$(4)$$

which shows that under translation each component of $\widehat{f}$ is simply premultiplied by an $e^{-i2\pi zk/n}$ factor.

The invariance of the spectrum is the result of the fact that in (2) these factors cancel:

$$q^z(k) = \left(e^{-i2\pi zk/n} \, \widehat{f}(k)\right)^* \cdot \left(e^{-i2\pi zk/n} \, f(k)\right) =$$
$$e^{i2\pi zk/n} \, \widehat{f}^*(k) \, e^{-i2\pi zk/n} \, f(k) = \widehat{f}^*(k) \cdot f(k) = q(k).$$

The spectrum is often used in signal processing applications as a translation invariant characterization of functions. Unfortunately, in computing the spectrum we lose all phase information: the spectrum only measures the energy in each band, not its phase relative to other bands.

The idea behind bispectral invariants is to move from (3) to the **triple correlation**

$$a(x_1, x_2) = \sum_{y=0}^{n-1} f^*(y - x_1) \, f^*(y - x_2) \, f(y).$$

Note that in some of the literature the triple correlation is defined slightly differently, and the above quantity would be $a^*(-x_1, -x_2)$. We deviate from this convention so as to make the formulae involved in the generalization to groups slightly more transparent. Again by the convolution theorem, the (2-dimensional) Fourier transform of this function is

$$b(k_1, k_2) = \widehat{f}^*(k_1) \, \widehat{f}^*(k_2) \, \widehat{f}(k_1 + k_2),$$

and this is what is called the **bispectrum** of f. Under translation b becomes

$$b^z(k_1, k_2) = e^{i2\pi z k_1/n} \widehat{f}^*(k_1) \cdot e^{i2\pi z k_2/n} \widehat{f}^*(k_2) \cdot$$
$$e^{-i2\pi z(k_1+k_2)/n} \widehat{f}(k_1 + k_2) = b(k_1, k_2),$$

so the bispectrum is invariant. The remarkable fact is that unlike the ordinary power spectrum, b is also sufficient to reconstruct the original signal up to translation. The bispectrum is widely used in signal processing as a lossless shift-invariant representation, and various algorithms have been devised to reconstruct f from b.

2.1 Bispectrum on groups

The "Euclidean" bispectrum introduced above would already be sufficient to construct translation invariant kernels. However, if we are to construct a kernel which is invariant to both translation and rotation, due to the intricate way in which these operations interact, we need to take a slightly more abstract viewpoint and re-examine what was said above from the point of view of group theory. While the concept of "Euclidean" bispectra is fairly well known in signal processing and computer vision, its generalization to non-commutative groups has attracted much less attention. The pioneering researcher in this field was R. Kakarala [3].

Recall that a group G is a set with a multiplication operation $\cdot : G \times G \to G$ obeying the following axioms:

G1 For any $x, y \in G$, $xy \in G$ (closure);

G2 For any $x, y, z \in G$, $(xy)z = x(yz)$ (associativity);

G3 There is a unique element of G denoted e and called the **identity** for which $ex = xe = x$ for any $x \in G$;

G4 For any $x \in G$ there is a corresponding element $x^{-1} \in G$ called the **inverse** of x, which satisfies $xx^{-1} = x^{-1}x = e$ for any $x \in G$.

Significantly, groups need not be commutative, i.e., xy need not equal yx. This is crucial for our present purposes since rigid planar motions don't commute.

Given a group G and a function $f : G \to \mathbb{C}$ to define the Fourier transform of f we need to introduce the concept of **group representations**. A representation is essentially a way of modeling the group operation by the multiplication of complex valued matrices. We say that $\rho : G \to \mathbb{C}^{d_\rho \times d_\rho}$ is a representation of G if

$$\rho(xy) = \rho(x)\,\rho(y)$$

for any $x, y \in G$. We also require $\rho(e) = I$. We say that d_ρ is the dimensionality of the representation. Note that $\rho(x^{-1}) = (\rho(x))^{-1}$.

There are some trivial ways of producing new representations from existing ones. For example, if ρ_1 is a representation of G, then for any invertible matrix T, so is $T^{-1}\rho_1(x)\,T$. These representations are clearly not substantially different, so they are called **equivalent**.

Another way that representations may be related is when a larger representation splits into smaller ones. We say that ρ is **reducible** if some invertible square matrix T can block diagonalize it in the form

$$T^{-1}\rho(x)\,T = \left(\begin{array}{c|c} \rho_1(x) & 0 \\ \hline 0 & \rho_2(x) \end{array} \right) \qquad x \in G$$

into a direct sum of smaller representations ρ_1 and ρ_2.

To develop the theory what are really important are the **irreducible** representations that cannot be reduced in this way. Given a group G there is a lot of interest in constructing a complete set of inequivalent representations for it. Such a set we will denote by $\mathcal{R}$. For a wide range of groups we can choose $\mathcal{R}$ to consist exclusively of unitary representations, so from now on we assume that $\rho(x^{-1}) = \rho(x)^\dagger$, where † denotes the conjugate transpose.

With these concepts of representation theory in hand, we return to (1) and note that the exponential factors appearing in the summation are nothing but representations (specifically, one-dimensional, irreducible representations) of the group formed by $\{0, 1, 2, \ldots, n-1\}$ with respect to addition modulo n. This suggests generalizing Fourier transformation to the non-commutative realm in the form

$$\widehat{f}(\rho) = \sum_{x \in G} f(x)\,\rho(x) \qquad \rho \in \mathcal{R}. \qquad (5)$$

Here and in the following the summation sign either denotes a discrete sum over the elements of a discrete group, or an integral (with respect to Haar measure) over a Lie group. Note that in contrast to (1), for general groups the components of $\widehat{f}$ are matrices and not scalars, and they are not indexed by the elements of G, but by its irreducible representations.

The generalized Fourier transform shares many important properties with its Euclidean counterpart, but most of these will not concern us here. What is important is that there is a natural concept of translation of functions on G defined by

$$f^z(x) = f(z^{-1}x) \qquad z \in G,$$

and that by the defining property of representations,

$$\widehat{f^z}(\rho) = \sum_{x \in G} f(z^{-1}x)\,\rho(x) =$$

$$\sum_{x \in G} f(z^{-1}x)\,\rho(z)\rho(z^{-1}x) =$$

$$\rho(z) \sum_{x \in G} f(x)\,\rho(x) = \rho(z)\widehat{f}(\rho)$$

in exact analogy with (4). In particular, by the unitarity of ρ, the generalized power spectrum $q(\rho) = \widehat{f}(\rho)^\dagger \widehat{f}(\rho)$ is again invariant to translation:

$$q^z(\rho) = \left(\rho(z)\widehat{f}(\rho)\right)^\dagger \left(\rho(z)\widehat{f}(\rho)\right) =$$
$$f(\rho)^\dagger \rho(z)^\dagger \rho(z) f(\rho)^\dagger = \widehat{f}(\rho)^\dagger \widehat{f}(\rho).$$

As in the classical case, the power spectrum does not uniquely determine f. The loss of information is related to the fact that the $q(\rho)$ matrices are by definition constrained to be positive definite, and again the power spectrum is insensitive to phase information in the sense that we may multiply any Fourier component by a different invertible matrix without affecting the power spectrum.

To construct the bispectrum we need to couple the different components of $\widehat{f}$, while at the same time retaining invariance. Consider tensor products $\widehat{f}(\rho_1) \otimes \widehat{f}(\rho_2)$, which transform according to

$$\widehat{f^z}(\rho_1) \otimes \widehat{f^z}(\rho_2) = (\rho_1(z) \otimes \rho_2(z)) \left(\widehat{f}(\rho_1) \otimes \widehat{f}(\rho_2)\right).$$

Now $\rho_1(z) \otimes \rho_2(z)$ is also a representation of G, but typically it is not irreducible. However, for wide classes of groups tensor product representations decompose into irreducibles in the form

$$\rho_1(z) \otimes \rho_2(z) = C \left[\bigoplus_\rho \rho(z) \right] C^\dagger.$$

Determining which set of irreducibles the direct sum ranges over (and with what multiplicities) and what the unitary matrix C should be is in general a highly non-trivial problem in representation theory. For now we assume that this so-called Clebsch-Gordan decomposition is known.

In this case we have a generalized bispectrum

$$b(\rho_1, \rho_2) = C^\dagger \left(\widehat{f}(\rho_1) \otimes \widehat{f}(\rho_2)\right)^\dagger C \bigoplus_\rho \widehat{f}(\rho), \quad (6)$$

and it will be translation invariant, $b^z(\rho_1, \rho_2) = b(\rho_1, \rho_2)$. What goes beyond a straightforward generalization of the classical results is the proof that for a wide range of groups, including all compact groups, if all $\widehat{f}(\rho)$ Fourier components are invertible matrices,

then b uniquely determines f up to translation. This is a highly technical result proved in [3], and in constrast to the commutative case, there might not be an algorithm for recovering f.

2.2 Homogeneous spaces

Before addressing the problem of image invariants, we need one more technical extension of the foregoing. We say that a group G **acts** on a space X, if for any $g \in G$ there is a mapping $T_g \colon X \to X$ such that if $g_2 g_1 = g_3$, then $T_{g_1}(T_{g_2}(x)) = T_{g_3}(x)$ for any $x \in X$. Now X is a **homogeneous space** of G if fixing any $x_0 \in X$, the set $T_g(x_0)$ ranges over the whole of X as g ranges over G. The classical example of a homogenous space, which will also be our choice for our image recognition problem, is the unit sphere S_2. The sphere is a homogeneous space of the three-dimensional rotation group SO(3): taking the North pole as x_0, a suitable rotation can move it to any point $x \in S_2$.

Fourier transformation generalizes naturally to functions $f \colon X \to \mathbb{C}$:

$$\widehat{f}(\rho) = \sum_{g \in G} f(T_g(x_0))\,\rho(g) \qquad \rho \in \mathcal{R},$$

as does the concept of translation, $f^g(x) = f(T_{g^{-1}}(x))$, and the bispectrum (6) remains invariant to such translations.

Note that except for the trivial case $X = G$, Fourier transforms on homogeneous spaces are naturally redundant: typically X is a much smaller space than G, yet a Fourier transform on X has the same number of components as a Fourier transform on the entire group. One manifestation of this fact is that we might find that some Fourier components are rank deficient no matter what $f \colon X \to \mathbb{C}$ we choose. While this destroys Kakarala's uniqueness result, in practice we often find that the bispectrum still furnishes a remarkably rich invariant representation of f. We remark that that invariance to right-translation $f^{(z)}(x) = f(xz^{-1})$ would be a different matter: there is a variant of the bispectrum which retains the uniqueness property in this case (Theorem 3.3.6 in [3]).

3 Bispectral invariants for images

After the abstract discussion of the previous section we now set out to construct concrete invariants for 2D monochrome images. We represent an image as an intensity function $h \colon \mathbb{R}^2 \to [0, 1]$ with support confined to a compact region of the plane, for example, the square $[-0.5, 0.5]^2$. The group that we would ideally like to be working with encompassing all translations and rotations is the Euclidean group $\text{ISO}^+(2)$ of rigid

body motions in the plane. $\mathbb{R}^2$ is a homogeneous space of $\mathrm{ISO}^+(2)$, so we could compute the $\mathrm{ISO}^+(2)$-Fourier transform of our image, and construct its bispectrum as described above.

The problem with this approach is that $\mathrm{ISO}^+(2)$ is not compact. Although it does belong to a class of exceptional groups to which Kakarala's uniqueness result does apply, its representation theory is complicated and computing the bispectrum is likely to be computationally very challenging. The main contribution of this paper is to show how the reduce the problem to rotations of the sphere. The rotation group SO(3) also happens to have the simplest and best known non-trivial Clebsch-Gordan decomposition. To make the exposition as elementary as possible, we derive the bispectral invariants from first principles, exploiting the simplifications afforded by this special case.

3.1 Projection onto the sphere

We begin by projecting our image h onto the unit sphere S_2. The simplest possible projection is to project parallel to the z-axis, formally

$$h \mapsto f, \qquad f(\theta, \phi) = h(r_{\mathbb{R}^2}, \theta_{\mathbb{R}^2}) = h(\tfrac{1}{a}\theta, \phi), \quad (7)$$

where $0 \le \theta \le \pi$ and $0 \le \phi < 2\pi$ are spherical polar coordinates, while $r_{\mathbb{R}^2} = \frac{1}{a}\theta$ and $\theta_{\mathbb{R}^2} = \phi$ are planar polars. The magnification parameter a we are free to choose between reasonable bounds as long as our image "fits" on the surface of the sphere. Inevitably, such a mapping does involve some distortion, particulary at the corners, as the image conforms to the curved surface of S_2. Reducing a decreases this distortion at the expense of reducing the surface area of the sphere actually occupied by the image, and hence increasing the computaional cost at the same effective resolution. In practice, even relatively large values of a (up to 1.5) do not hurt performance. Apart from the inevitable finite bandwidth cutoff, this is the only approximation involved in our method.

To numerically represent f we use **spherical harmonics**

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

where $l = 0, 1, 2, \ldots$; $m = -l, -l+1, \ldots, l$ and P_l^m are the associated Legendre polynomials. Recall that the spherical harmonics are the eignefunctions of the Laplace operator on S_2 (with eigenvalue $-l^2$), and they form an orthonormal basis for $L_2(S_2)$, thus we can represent f as

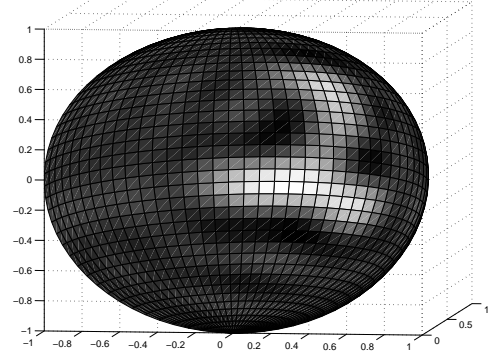$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \widehat{f}_{l,m} \, Y_l^m(\theta, \phi) \qquad (8)$$

Figure 1: A NIST handwritten digit projected onto the sphere. The band-limit is $L = 15$. Note that there is a minimal amount of "ringing".

where $\widehat{f}_{l,m} = \langle f, Y_l^m \rangle$ and $\langle \cdot, \cdot \rangle$ is the inner product

$$\langle f, g \rangle = \int_0^\pi \int_0^{2\pi} f^*(\theta, \phi) \, g(\theta, \phi) \cos\theta \, d\phi \, d\theta.$$

We denote by $\widehat{\boldsymbol{f}}_l$ the vector $(\widehat{f}_{l,-l}, \widehat{f}_{l,-l+1}, \ldots, \widehat{f}_{l,l})$.

Viewing S_2 as a homogeneous space of SO(3), the $\{\widehat{f}_{l,m}\}$ are the Fourier coefficients of $f \colon S_2 \to \mathbb{C}$ as defined in the previous section. However, in this special case they do not form matrices, only vectors: if we formally computed (2.2), we would find that only the first column of each matrix is non-zero (see also [1]). This will make the computational burden significantly lighter.

In a computational setting we must truncate (8) at some finite L, preferably so as to match the resolution of our original image. In general, the spherical representation of an image requires more storage than the original pixmap representation only to the extent that the image only occupies a fraction of the surface of the sphere.

For a $[0, 1]$-valued bitmap matrix M, the mapping (7) leads to

$$\widehat{f}_{l,m} = \sum_{i,j=1}^{n} M_{i,j} Y_l^m(\theta, \phi), \qquad (9)$$

where $\theta = a\sqrt{x^2 + y^2}$,

$$\phi = \begin{cases} \arctan(y/x) & \text{if } y > 0 \\ 2\pi - \arctan(y/x) & \text{if } y < 0 \end{cases},$$

and $(x, y) = \left(\frac{i - 1/2}{N} - 0.5, \; \frac{j - 1/2}{N} - 0.5 \right)$.

Just as the isometry group of $\mathbb{R}^2$ is $\mathrm{ISO}^+(2)$, the isometry group of S_2 is SO(3), the group of rotations of $\mathbb{R}^3$ about the origin. It is easy to visualize that given the mapping (7), locally, around the north pole, there is a
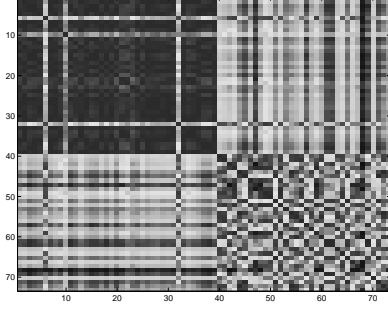
Figure 2: The inner product matrix between the bispectrum representation of the "0" and "1" digits from the first 300 translated and rotated NIST characters. The block structure reflects that the intra-class inner products are higher than the inter-class products.

one-to one corresponence between the action of SO(3) on functions on the sphere and of $\text{ISO}^+(2)$ on the corresponding functions on the plane. In other words, any rigid motion of an image in the plane can be imitated by a 3D rotation of the corresponding function on S_2. Rotations of the image around the center of the image correspond to rotations of the sphere about the z axis (pole to pole), while translations correspond to rotations around the x and y axes. Exploiting this fact, we proceed by computing the bispectral invariants of f with respect to SO(3) and let these be our translation and rotation invariant features.

3.2 An SO(3)-invariant kernel on $L_2(S_2)$

To construct the SO(3)-invariant features, we examine how SO(3) acts on individual spherical harmonics. Since $\{Y_l^m\}_{m=-l,\dots,l}$ span the space of eigenvectors of the Laplace operator with eigenvalue $-l^2$, and since the Laplace operator is rotationally invariant, under the action of a rotation $R \in$ SO(3), Y_l^m must transform into a linear combination $R(Y_l^m) = \sum_{m'=-l}^{l} a_m Y_l^{m'}$ of other spherical harmonics of the same order l.

For a general function $f \in L_2(S_2)$, under a rotation $R \in$ SO(3) the Fourier coefficients transform according to

$$
\begin{pmatrix} \widehat{f}'_{l,-l} \\ \vdots \\ \widehat{f}'_{l,l} \end{pmatrix} = D^{(l)}(R) \begin{pmatrix} \widehat{f}_{l,-l} \\ \vdots \\ \widehat{f}_{l,l} \end{pmatrix}, \quad (10)
$$

where $D^{(l)}(R)$ are $(2l+1) \times (2l+1)$ dimensional matrices. In fact, $D^{(0)}, D^{(1)}, \dots$ are exactly the (complex-valued) irreducible representations of SO(3).

It is possible to show that the $D^{(l)}$ are unitary repre-

sentations, hence the polynomials

$$
p_l = \sum_{m=-l}^{l} \left| \widehat{f}_{l,m} \right|^2 = \widehat{\boldsymbol{f}}_l^\dagger \cdot \widehat{\boldsymbol{f}}_l = \left(\widehat{f}_{l,-l}^*, \dots, \widehat{f}_{l,l}^* \right) \cdot \begin{pmatrix} \widehat{f}_{l,-l} \\ \vdots \\ \widehat{f}_{l,l} \end{pmatrix}
$$

transform according to

$$
\begin{aligned}
p_l &\mapsto \left(D^{(l)}(R)\, \widehat{\boldsymbol{f}}_l \right)^\dagger \cdot \left(D^{(l)}(R)\, \widehat{\boldsymbol{f}}_l \right) = \\
&\qquad \widehat{\boldsymbol{f}}_l^\dagger \left(D^{(l)}(R) \right)^\dagger \left(D^{(l)}(R) \right) \widehat{\boldsymbol{f}}_l = \widehat{\boldsymbol{f}}_l^\dagger \cdot \widehat{\boldsymbol{f}}_l,
\end{aligned}
$$

i.e., they are invariant. This is the power spectrum, as defined in Section 2.1. As before, this is an invariant, but very impoverished representation of f.

The bispectrum is derived by considering the $(2l_1 + 1)(2l_2+1)$-dimensional tensor product vectors $\widehat{\boldsymbol{f}}_{l_1} \otimes \widehat{\boldsymbol{f}}_{l_2}$, which transform according to

$$
\widehat{\boldsymbol{f}}_{l_1} \otimes \widehat{\boldsymbol{f}}_{l_2} \mapsto \left(D^{(l_1)}(R) \otimes D^{(l_2)}(R) \right) \cdot \left(\widehat{\boldsymbol{f}}_{l_1} \otimes \widehat{\boldsymbol{f}}_{l_2} \right). \quad (11)
$$

The representation theory of SO(3) is well developed, in particular, it is well known that the tensor product representations decompose in the form

$$
\begin{aligned}
D^{(l_1)}(R) \otimes D^{(l_2)}(R) = \\
\left(C^{l_1,l_2} \right)^\dagger \left[\bigoplus_{l=|l_1-l_2|}^{l_1+l_2} D^{(l)}(R) \right] C^{l_1,l_2}.
\end{aligned}
$$

Here C^{l_1,l_2} is a $((2l_1+1)(2l_2+1)) \times ((2l_1+1)(2l_2+1))$-element unitary matrix, with rows labeled by the pair (l,m) and columns labeled by the pair (m_1, m_2). The matrix elements $C^{l_1,l_2,l}_{m_1,m_2,m} = \left[C^{l_1,l_2} \right]_{(l,m),(m_1,m_2)}$ are called **Clebsch-Gordan coefficients**, and are implemented in most computational algebra packages. Our notation is redundant in that it is possible to show that $C^{l_1,l_2,l}_{m_1,m_2,m}$ vanishes unless $m_1 + m_2 = m$, hence we only need to worry about the coefficients $C^{l_1,l_2,l}_{m_1,m-m_1,m}$.

Thus, under rotation $C^{l_1,l_2}\left(\widehat{\boldsymbol{f}}_{l_1} \otimes \widehat{\boldsymbol{f}}_{l_2} \right)$ transforms according to

$$
C^{l_1,l_2}\left(\widehat{\boldsymbol{f}}_{l_1} \otimes \widehat{\boldsymbol{f}}_{l_2} \right) \mapsto \left[\bigoplus_{l=|l_1-l_2|}^{l_1+l_2} D^{(l)}(R) \right] C^{l_1,l_2}\left(\widehat{\boldsymbol{f}}_{l_1} \otimes \widehat{\boldsymbol{f}}_{l_2} \right).
$$

$$ (12) $$

Writing $C^{l_1,l_2}\left(\widehat{\boldsymbol{f}}_{l_1} \otimes \widehat{\boldsymbol{f}}_{l_2} \right) = \bigoplus_{l=|l_1-l_2|}^{l_1+l_2} \widehat{\boldsymbol{g}}_{l_1,l_2,l}$, where

$$
\left[\widehat{\boldsymbol{g}}_{l_1,l_2,l} \right]_m = \sum_{m_1=-l_1}^{l_1} C^{l_1,l_2,l}_{m_1,m-m_1,m} \, \widehat{f}_{l_1,m_1} \, \widehat{f}_{l_2,m-m_1},
$$

$\widehat{\boldsymbol{g}}_{l_1,l_2,l}$ transforms according to

$$
\widehat{\boldsymbol{g}}_{l_1,l_2,l} \mapsto D^{(l)}(R)\, \widehat{\boldsymbol{g}}_{l_1,l_2,l}.
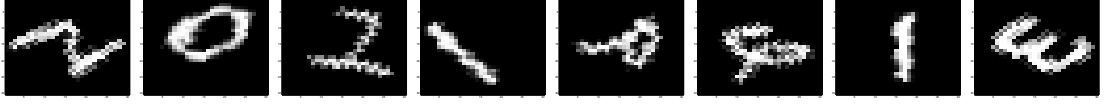$$

Figure 3: The first few rotated and translated NIST characters.

By the same argument as for the power spectrum, this gives rise to the cubic invariants

$$p_{l_1,l_2,l} = \widehat{\boldsymbol{g}}^{\dagger}_{l_1,l_2,l} \cdot \widehat{\boldsymbol{f}}_l =$$
$$\sum_{m=-l}^{l} \sum_{m_1=-l_1}^{l_1} C^{l_1,l_2,l}_{m_1,m-m_1,m} \widehat{f}^{*}_{l_1,m_1} \widehat{f}^{*}_{l_2,m-m_1} \widehat{f}_{l,m}. \quad (13)$$

Up to unitary transformation, these invariants are equivalent to the non-vanishing matrix elements of the abstract bispectrum (as already derived in [3] and [1]). Any kernel built from the bispectrum using (13) as features will be invariant to translation and rotation.

3.3 Computational considerations

The algorithmic implementation of (13) is

$$p_{l_1,l_2,l} = \sum_{m=-l}^{l} \widehat{f}_{l,m} \times$$
$$\times \sum_{m_1=\max(-l_1,m-l_2)}^{\min(l_1,m+l_2)} C^{l_1,l_2,l}_{m_1,m-m_1,m} \widehat{f}^{*}_{l_1,m_1} \widehat{f}^{*}_{l_2,m-m_1},$$

which gives $O(L^3)$ invariant features to build the kernel from. The features can be precomputed as a data processing step before any learning actually takes place. Typically, L will scale linearly with the linear dimension w of the input image in pixels, so the bispectrum inflates the data at a rate of $u^{3/2}$, where u is the original storage size of a single image.

Projecting onto the sphere is a linear map and its coefficients can be precomputed, so the cost of that operations scales with $w^2 L^2 \propto u^2$. Finally, computing the bispectrum itself scales with $L^5 \propto u^{5/2}$. On the desktop PC used to prepare the data for the experiments, processing each 30×30 pixel image took approximately 100ms for $L = 15$.

4 Experiments

We conducted experiments on randomly translated and rotated versions of hand-written digits from the well known NIST dataset [6]. The original images are size 28×28, but most of them only occupy a fraction of the image patch. The characters are rotated by a random angle between 0 and 2π, clipped, and embedded at a random position in a 30×30 patch (fig. 3).

We trained 2-class SVMs for all possible pairs of digits. As a baseline we used SVMs with linear and Gaussian RBF kernels on the original 900-dimensional pixel intensity vector. We compared this to similar linear and Gaussian RBF SVMs ran on the bispectrum features. We used $L = 15$, which is a relatively low resolution for images of this size. The magnification parameter was set to $a = 2$.

Our experimental procedure consisted of using cross-validation to set the regularization parameter C and the kernel width σ independently for each each learning task: digit d_1 vs. digit d_2. We used 10-fold cross validation to set the parameters for the linear kernels, but to save time only 3-fold cross validation for the Gaussian kernels. Testing and training was conducted on the relevant digits from the second one thousand images in the NIST dataset. The results we report are averages and standard deviations of error for 10 random even splits of this data. Since there are on average 100 digits of each type amongst the 1000 images in the data, our average training set and test set consisted of just 50 digits of each class. Given that the images also suffered random translations and rotations this is an artificially difficult learning problem.

The results are shown in table 3.3 for the linear kernel and in table 3.3 for the RBF kernel. The two sets of results are very similar. In both cases the bispectrum features far outperform the baseline bitmap representation. Indeed, it seems that in many cases the baseline cannot do better than what is essentially random guessing. In contrast, the bispectrum can effectively discriminate even in the hard cases such as 8 vs. 9 and reaches almost 100% accuracy on the easy cases such as 0 vs. 1. Surprisingly, to some extent the bispectrum can even discriminate between 6 and 9, which in some fonts are exact rotated versions of each other. However, in handwriting, 9's often have a straight leg and/or a protrusion at the top where right handed scribes reverse the direction of the pen.

The results make it clear that the bispectrum features are able to capture position and orientation invariant characteristics of handwritten figures. We did not compare our algorithm against other image kernels due to time constraints. However, short of a handwriting-specific algorithm which extracts exlicit landmarks we do not expect other methods to yield a comparable degree of position and rotation invariance.

	1	2	3	4	5	6	7	8	9
0	0.77(0.41) 17.12(3.67)	6.22(2.41) 33.87(3.59)	5.09(1.54) 42.06(3.59)	5.03(1.07) 30.64(2.53)	2.90(1.53) 37.82(3.51)	4.11(2.39) 31.42(5.85)	2.73(1.11) 29.36(3.83)	4.98(1.64) 42.58(4.33)	5.86(2.88) 27.61(3.16)
1		0.68(0.81) 30.78(2.90)	0.39(0.98) 29.34(4.50)	3.07(1.30) 34.96(3.41)	0.00(0.00) 30.66(2.85)	1.37(0.88) 34.46(4.47)	1.77(1.48) 38.32(4.05)	2.68(2.02) 24.60(2.57)	1.02(1.00) 34.78(3.57)
2			15.89(5.79) 49.06(4.18)	15.82(3.22) 47.12(4.72)	8.06(3.60) 45.20(4.26)	9.64(2.00) 51.44(5.21)	11.11(2.29) 47.20(5.54)	9.26(1.63) 47.44(6.23)	10.55(2.95) 46.70(2.95)
3				4.81(1.68) 44.64(3.03)	16.42(5.69) 49.07(4.81)	7.54(2.75) 49.38(5.26)	4.00(1.13) 44.74(4.42)	10.70(3.79) 50.37(4.21)	7.66(3.01) 47.60(5.55)
4					6.26(1.90) 40.08(6.67)	10.94(4.09) 50.11(5.26)	14.95(2.89) 45.30(3.30)	6.27(3.57) 46.26(2.63)	16.95(1.84) 49.82(4.68)
5						14.63(2.42) 50.00(4.02)	5.31(2.27) 41.70(4.09)	6.62(2.72) 44.63(3.31)	6.84(2.23) 46.01(4.37)
6							7.68(4.05) 48.19(4.10)	9.00(2.93) 46.13(5.82)	20.15(3.62) 53.75(2.69)
7								3.50(2.28) 41.16(5.18)	8.06(3.49) 53.21(5.01)
8									9.43(2.14) 45.13(2.87)

Table 1: Classification error in percent for each pair of digits for the linear kernels. The performance of the bispectrum-based classifier is shown on top, and the baseline on bottom; standard errors are in parentheses.

	1	2	3	4	5	6	7	8	9
0	0.80(0.42) 12.50(3.60)	5.06(1.52) 26.30(4.32)	4.78(1.08) 33.72(4.58)	3.35(1.69) 32.45(12.63)	3.90(2.25) 29.52(3.99)	3.07(1.77) 23.51(4.93)	4.48(1.39) 24.96(3.73)	3.74(2.23) 29.99(4.20)	6.34(2.57) 19.16(2.65)
1		0.99(0.48) 27.29(4.00)	0.00(0.00) 22.61(8.82)	2.48(0.97) 33.98(9.44)	0.21(0.45) 30.86(9.99)	1.35(0.43) 28.52(9.47)	1.22(1.09) 32.12(6.34)	0.52(0.55) 20.16(2.93)	3.05(0.88) 28.01(4.56)
2			14.68(4.60) 47.75(3.46)	13.20(2.56) 45.26(5.11)	8.83(4.22) 50.09(4.78)	8.89(3.09) 45.63(5.49)	12.73(3.39) 43.84(4.38)	12.14(2.27) 44.02(3.14)	10.34(2.51) 45.95(4.84)
3				5.12(2.35) 43.07(9.05)	16.88(2.73) 52.53(3.39)	6.98(3.46) 45.86(5.27)	3.50(1.48) 41.90(4.09)	10.21(3.89) 46.00(4.97)	5.08(1.50) 44.87(3.91)
4					5.75(1.22) 39.21(4.29)	10.67(1.47) 46.82(5.32)	13.92(2.63) 46.73(6.47)	6.45(2.26) 42.29(4.44)	12.09(2.47) 52.73(3.65)
5						16.56(1.66) 47.04(4.21)	6.26(1.54) 46.39(3.41)	6.23(3.05) 41.63(3.29)	7.07(2.93) 43.23(2.46)
6							9.30(3.33) 40.43(5.16)	6.16(2.30) 41.19(4.47)	21.37(3.81) 50.73(4.31)
7								4.68(2.30) 37.33(2.21)	8.81(2.81) 46.22(4.13)
8									10.06(2.04) 44.06(3.93)

Table 2: Classification error in percent for each pair of digits for the Gaussian RBF kernels. The performance of the bispectrum-based classifier is shown on top, and the baseline on bottom, standard errors are in parentheses.

5   Conclusions

We presented an application of the theory of bispectra on non-commutative groups to constructing a novel system of translationally and rotationally invariant features for images. The method hinges on a projection from the plane to the sphere, reducing the problem of invariance to the action of the non-compact Euclidean group to that of the compact and computationally tractable three dimensional rotations group.

Our method may be used as a pre-processing step for learning algorithms, in particular, kernel-based discriminative algorithms. Computational requirements scale with $u^{5/2}$ and memory requirements with $u^{3/2}$ where u is the size of the original image (in pixels).

Experimental results on an optical character recognition problem indicate that the method is surprisingly powerful "out of the box". Time constraints prevented us from conducting more extensive experiments on larger images (entire scenes), multicolor images, etc., but we expect our algorithm to remain viable over a range of tasks.

Finally, we believe that the general concept of bispectra ought to be of interest to the machine learning community as it moves towards addressing learning tasks on more and more intricately structured data. This motivated the general discussion of the bispectrum concept in the first half of this paper.

6   Acknowledgments

References

[1] Dennis M. Healy, Daniel N. Rockmore, and Sean S. B. Moore. FFTs for the 2-sphere. Improvements and variations. Technical Report PCS-TR96-292, Dartmouth College, 1996.

[2] Janne Heikkilä. A new class of shift-invariant operators. Technical report, Machine Vison Group, Dept. of Electrical and Information Engineering, U. of Oulu, Finland, 2004.

[3] R Kakarala. *Triple corelation on groups.* PhD thesis, Department of Mathematics, UC Irvine, 1992.

[4] R. Kakarala. A group theoretic approach to the triple correlation. In *IEEE Workshop on higher order statistics*, pages 28–32, 1993.

[5] R. Kondor and T. Jebara. A kernel between sets of vectors. In *Proceedings of the ICML*, 2003.

[6] Yann LeCun and Corinna Cortes. The nist dataset provided at `http://yann.lecun.com/db/mnist/`.

[7] M Michaelis and G Sommer. A Lie group-approach to steerable filters. *Pattern Recognition Letters*, 16(11), 1995.

[8] Brian M. Sadler and Georgios B. Giannakis. Shift- and rotation-invariant object recognition using the bispectrum. *Journal of the Optical Society of America, A*, 9(1):57–69, 1992.

[9] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond.* MIT Press, Cambridge, MA, 2001.