

SHAPE REPRESENTATION VIA ELEMENTARY SYMMETRIC POLYNOMIALS: A COMPLETE INVARIANT INSPIRED BY THE BISPECTRUM

Renato M. P. Negrinho, Pedro M. Q. Aguiar

Institute for Systems and Robotics / Instituto Superior Técnico, Lisboa, Portugal
renato.negrinho@ist.utl.pt, aguiar@isr.ist.utl.pt

ABSTRACT

We address the representation of two-dimensional shape in its most general form, *i.e.*, arbitrary sets of points, that may arise in multiple situations, *e.g.*, sparse sets of specific landmarks, or dense sets of image edge points. Our goal are recognition tasks, where the key is balancing two contradicting demands: shapes that differ by rigid transformations or point re-labeling should have the same representation (*invariance*) but geometrically distinct shapes should have different representations (*completeness*). In the paper, we introduce a new shape representation that marries properties of the elementary symmetric polynomials and the bispectrum. Like the power spectrum, the bispectrum is insensitive to signal shifts; however, unlike the power spectrum, the bispectrum is complete. The elementary symmetric polynomials are complete and invariant to variable re-labeling. We show that the elementary symmetric polynomials of the shape points depend on the shape orientation in a way that enables interpreting them in the frequency domain and building from them a bispectrum. The result is a shape representation that is complete and invariant to rigid transformations and point-relabeling. The paper also reports experiments that illustrate the proved properties.

Index Terms— Shape representation, Complete invariant, Bispectrum, Elementary symmetric polynomials, Shape recognition

1. INTRODUCTION

Representing two-dimensional shape remains a challenge. Although connected regions can be represented by a one-dimensional contour, which is easier to code, see, *e.g.*, [1], this is not the case of general shapes, *i.e.*, arbitrary sets of points in the plane. The statistical theory of shape [2] addresses this problem in situations where the points are labeled (usually in small number, denoted by landmarks). However, the problem remains for reasonably large sets of points without labels or natural ordering, *e.g.*, those arising from automatic edge/corner/interest-point detection.

The motivation for representing shapes comes from the need to compare them in recognition applications. Comparing two collections of points is difficult, since they are related by unknown geometric transformations (due do different position, orientation, and size) and permutation (due to the absence of labels for the points). Although translation and scale are easily taken care of through normalization, this is not the case of rotation and permutation, whose simultaneous estimation leads to a non-convex problem. Iterative methods such as the Iterative Closest Point (ICP) [3] or its probabilistic versions based on Expectation-Maximization (EM), *e.g.*, [4]

tackle this problem but suffer from the usual sensitivity to the initialization, exhibiting uncertain convergence. When the relative orientation of the shapes to compare is known, the estimation of the permutation relating the point sets can be casted into a convex optimization problem [5]. However, normalizing a point set in what respects to rotation is harder than it could seem at first sight. In fact, although theoretically sustained moment-based methods have been proposed (see [6] and the references therein), degenerate cases have been successively identified, showing that these methods can be sensitive to the noise and motivating subsequent research, *e.g.*, [7].

The difficulties summarized in the previous paragraph but also the fact that modern machine learning algorithms require more than the capability of comparing pairs of shapes, motivates the search for a representation that enable shapes to be treated as points in an abstract space, where machine learning algorithms can be applied. Naturally, the representation must be *invariant* to the geometric transformations and point permutation but also *complete*, in the sense of fully describing the underlying shape, *i.e.*, shapes that only differ by a rigid geometric transformation or point re-labeling are mapped to the same point in the abstract shape space while really distinct shapes are mapped to different points in that space. Reference [8] proposes a permutation-invariant representation that is also complete but it is not rotation-invariant, requiring pairwise alignment. Moment-based representations of image patterns have been used since the sixties due to their geometric invariance properties but their completeness only recently have been focus of attention [6].

In this paper we introduce a new shape representation that is complete and invariant with respect to (w.r.t.) geometric transformations. The representation is based on the so-called elementary symmetric polynomials of the shape points that are invariant w.r.t. shape point permutation and unambiguously determine the shape. To obtain complete invariance w.r.t. shape rotation, we draw inspiration from the bispectrum. While the power spectrum of a signal is insensitive to signal shifts but does not uniquely determine the underlying signal, its bispectrum inherits the invariance and determines the signal, up to a shift. We show that shape rotation affects the elementary symmetric polynomials in a similar way as a signal shift affects its Fourier transform. Based on this property, we propose a shape representation that consists in a bispectrum computed from the elementary symmetric polynomials, being then complete and invariant w.r.t. point permutation and shape orientation (translation and scale are taken care off through normalization).

The complete invariance of the bispectrum w.r.t. shifts has been used in image processing but, to the best of our knowledge, not to represent arbitrary sets of points. For example, in [9] image rotations are transformed into shifts in the polar domain and [10] uses the bispectrum of one-dimensional image projections (Radon transform). The completeness of bispectrum even inspired other authors to extend its applicability beyond commutative groups [11, 12].

This work was partially supported by *Fundação para a Ciência e a Tecnologia*, under grant [PEst-OE/EEI/LA0009/2011].

2. THE BISPECTRUM

The power spectrum of a finite-length discrete-time signal $x(n)$, $0 \leq n \leq N - 1$, is the discrete Fourier transform (DFT) of its autocorrelation, which is related to X , the DFT of x , by [13]:

$$P_x(k) = X(k)X^*(k) = |X(k)|^2, \quad 0 \leq k \leq N - 1.$$

If y is a circular shift of x , $y(n) = x((n - m) \bmod N)$, its DFT is

$$Y(k) = X(k)e^{-j\frac{2\pi}{N}km} \quad (1)$$

and its power spectrum is $P_y(k) = |Y(k)|^2 = |X(k)|^2 = P_x(k)$. The power spectrum is thus invariant to circular shifts. However, it is not complete, *i.e.*, signals that are not related by a shift may also have the same power spectrum, as illustrated in Fig. 1.

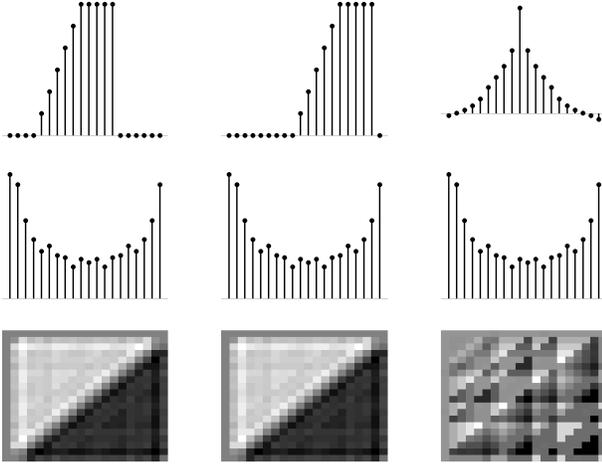


Fig. 1. Top: signals $x(n)$, $y(n) = x((n - 5) \bmod 20)$, and $z(n)$. Middle: P_x , P_y , and P_z , illustrate the invariance but also the incompleteness of the power spectrum. Bottom: $\arg B_x$, $\arg B_y$, and $\arg B_z$, illustrate that the bispectrum is shift-invariant and complete.

The bispectrum of x is the two-dimensional DFT of its triple-correlation, which, in terms of X , is given by

$$B_x(k, l) = X(k)X(l)X^*((k + l) \bmod N), \quad 0 \leq k, l \leq N - 1.$$

The shift-invariance of the bispectrum is established from the definition above and property (1) with simple manipulations:

$$\begin{aligned} B_y(k, l) &= Y(k)Y(l)Y^*((k + l) \bmod N) \\ &= X(k)e^{-j\frac{2\pi}{N}km} X(l)e^{-j\frac{2\pi}{N}lm} X^*((k + l) \bmod N)e^{j\frac{2\pi}{N}(k+l)m} \\ &= X(k)X(l)X^*((k + l) \bmod N) = B_x(k, l). \end{aligned}$$

Unlike the power spectrum, the bispectrum is complete, *i.e.*, it determines the underlying signal (obviously, up to a shift). Several proofs of this property, for increasingly classes of signals, were published since the eighties, *e.g.*, [14]. That shifted versions of a signal share a common bispectrum while distinct signals have different bispectrums, as desired in recognition, is also illustrated in Fig. 1, where, for compactness, only the arguments of the bispectrums are shown.

3. THE ELEMENTARY SYMMETRIC POLYNOMIALS

Given a set of N complex variables, z_1, z_2, \dots, z_N , the corresponding elementary symmetric polynomial of degree k , $C_z(k)$ is the sum of all possible products of k of them [15]. Formally, we define $I_k = \{(i_1, \dots, i_N) : i_n \in \{0, 1\}, 1 \leq n \leq N, \sum_{n=1}^N i_n = k\}$ and

$$C_z(k) = \sum_{(i_1, \dots, i_N) \in I_k} z_1^{i_1} z_2^{i_2} \dots z_N^{i_N}, \quad 0 \leq k \leq N. \quad (2)$$

A key characteristic of C_z is that it specifies the coefficients of the univariate monic polynomial $p(t)$ with roots z_1, \dots, z_N :

$$p(t) = \prod_{n=1}^N (t - z_n) = \sum_{k=0}^N (-1)^k C_z(k) t^{N-k}.$$

Obviously, the polynomial $p(t)$ does not depend on the order of the roots, which makes clear that C_z is invariant to permutations of the variables z_1, \dots, z_N , motivating the designation of *symmetric*, and that the set $\{C_z(k), 0 \leq k \leq N\}$ determines those variables, up to re-ordering.

4. THE PROPOSED SHAPE REPRESENTATION

Our goal is the representation of fully general two-dimensional shapes, *i.e.*, arbitrary collections of points in the plane. Two shapes should have the same representation if and only if the corresponding collections of points are related by arbitrary re-ordering and geometric transformations (translation, rotation, and scaling factor).

4.1. Translation and scale invariance

Let the complex variables w_1, \dots, w_N collect the plane coordinates of the shape points ($w_n = x_n + jy_n$). Translation and scale invariance are easily obtained through normalization, *i.e.*, by translating and scaling the original shape in order to produce a prototype whose centroid is at the origin and “power” is unitary. Formally, we obtain the normalized shape points, z_1, \dots, z_N , by pre-processing w_1, \dots, w_N , according to

$$z_n = \frac{w_n - \bar{w}}{\sqrt{p(w)}},$$

where $\bar{w} \in \mathbb{C}$ and $p(\cdot) \in \mathbb{R}^+$ denote centroid and power, defined as

$$\bar{w} = \frac{1}{N} \sum_{n=1}^N w_n, \quad p(w) = \frac{1}{N} \sum_{n=1}^N |w_n - \bar{w}|^2. \quad (3)$$

It is trivial to verify that $\bar{z} = 0$ and $p(z) = 1$. It is also simple to conclude that z represents w in the desired way, *i.e.*, that an arbitrary translation and scale of w leads to the same z and that z_1, \dots, z_N unambiguously determine the original shape points w_1, \dots, w_N , up to translation and scale.

4.2. Permutation invariance

To obtain invariance to the order by which the shape points are specified, we use the elementary symmetric polynomials of the normalized shape. In fact, as referred above, $\{C_z(k), 0 \leq k \leq N\}$, given by (2), forms a representation of the points z_1, \dots, z_N , which is invariant to point permutation and complete. This way, as desired, two shapes will have the same $\{C_z(k), 0 \leq k \leq N\}$ if and only if they differ by an arbitrary re-ordering.

We compute the sum in (2) by using an efficient recursive procedure, where the elementary symmetric polynomials for N points, $\{C_z^{(N)}(k)\}$, are obtained from those for $N - 1$ points through

$$C_z^{(N)}(k) = C_z^{(N-1)}(k) + z_N C_z^{(N-1)}(k-1).$$

This recursion is initialized with $C_z^{(1)}(0) = 1, C_z^{(1)}(1) = z_1$ but note that, regardless of the number of points, we have always $C_z(0) = 1$ and, for normalized shapes, $C_z(1) = 0$.

4.3. Rotation invariance

To deal with arbitrary orientations, we start by studying how the rotation of a shape is propagated to its elementary symmetric polynomials. Consider a shape z , with elementary symmetric polynomials C_z , and another shape s that only differs from z through a rotation of some angle θ , *i.e.*, $s_n = z_n e^{j\theta}, 1 \leq n \leq N$. The elementary symmetric polynomial of degree k for the rotated shape s is

$$\begin{aligned} C_s(k) &= \sum_{(i_1, \dots, i_N) \in I_k} s_1^{i_1} s_2^{i_2} \dots s_N^{i_N} \\ &= \sum_{(i_1, \dots, i_N) \in I_k} z_1^{i_1} e^{j\theta i_1} z_2^{i_2} e^{j\theta i_2} \dots z_N^{i_N} e^{j\theta i_N} \\ &= \sum_{(i_1, \dots, i_N) \in I_k} z_1^{i_1} \dots z_N^{i_N} e^{j\theta \sum_{n=1}^N i_n} \\ &= C_z(k) e^{jk\theta}, \end{aligned} \quad (4)$$

where we used the fact that $\sum_{n=1}^N i_n = k$, from the definition of I_k .

Comparing expressions (1) and (4), we see that shape rotations affect elementary symmetric polynomials in the same way as signal shifts affect DFT coefficients. Thus, a rotation-invariant complete shape representation may be derived in a similar way as the bispectrum lead to a shift-invariant complete signal representation. In particular, if we only care with rotation angles integer multiples of $2\pi/(N+1)$, *i.e.*, $\theta = m2\pi/(N+1), m \in \mathbb{Z}$, we get $C_s(k) = C_z(k) \exp(jkm2\pi/(N+1))$, and the parallel with discrete-time shifts renders exact. In this case, the shape z can thus be represented by a bispectrum computed from $\{C_z(k), 0 \leq k \leq N\}$ exactly as it was computed from $\{X(k), 0 \leq k \leq N-1\}$ in Section 2. Naturally, the invariance of this representation does not hold for other rotation angles. This is illustrated in Fig. 2 though the evolution of the magnitude of the difference between the representation of a shape z and the one of its rotated version s , as a function of the rotation angle θ : since $N = 4$, the error is zero for $\theta = m2\pi/5$.

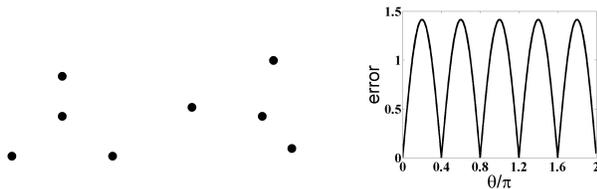


Fig. 2. Left: shape z . Middle: a rotated version s . Right: Frobenius norm of the difference between the matrices containing the DFT-based representations of z and s , as a function of the rotation angle.

The rotation-invariance just discussed may be adequate for particular cases, *e.g.*, when dealing with tiles (from a panel) whose orientation angles are restricted to integer multiples of a given angle.

However, we obtain a representation able to deal with arbitrary orientations by using the two following facts: i) unlike for the DFT analogy, expression (4) holds for an arbitrary rotation angle; and ii) a subset of the bispectral coefficients is invariant to arbitrary rotations. In summary, we propose to represent the shape z by the subset of bispectral coefficients $B_z(k, l)$ (of the elementary symmetric polynomials C_z) whose indexes satisfy $k + l \leq N$, which are

$$B_z(k, l) = C_z(k) C_z(l) C_z^*(k+l), \quad 0 \leq k, l \leq N, \quad k+l \leq N.$$

It is simple to verify that this representation is rotation-invariant:

$$\begin{aligned} B_s(k, l) &= C_s(k) C_s(l) C_s^*(k+l) \\ &= C_z(k) e^{jk\theta} C_z(l) e^{jl\theta} C_z^*(k+l) e^{-j(k+l)\theta} \\ &= C_z(k) C_z(l) C_z^*(k+l) \\ &= B_z(k, l). \end{aligned}$$

Note that the condition $k + l \leq N$ is key to avoid the wrap-around (*i.e.*, the mod) of the DFT-based representation that restricted its invariance to angles $\theta = m2\pi/(N+1)$, as illustrated in Fig. 2. The reader may wonder if the completeness of the bispectrum is lost by working only with a subset of its entries. However, it has been shown that the complete bispectrum is highly redundant and it is in fact possible, by following a strategy similar to the one in [9], to recover the underlying signal from the triangular subset $B_z(k, l) : k + l \leq N$. The example in Fig. 3 illustrates the invariance and completeness of the proposed shape representation.

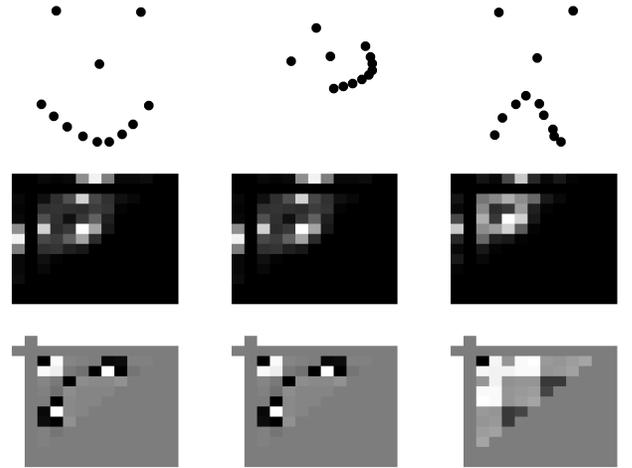


Fig. 3. Top: shapes z and s , related by a geometric transformation, and a distinct shape w . Middle and bottom: the corresponding representations ($|B_z|, |B_s|$, and $|B_w|$ in the middle and $\arg B_z, \arg B_s$, and $\arg B_w$ in the bottom), illustrating that the proposed representation is simultaneously invariant and complete ($B_z = B_s \neq B_w$).

5. ON THE NORMALIZATION OF THE POLYNOMIALS

The derivations of the previous section remain valid if the elementary symmetric polynomials are normalized according to a fixed procedure, *i.e.*, if $C_z(k)$ is multiplied by an arbitrary real positive number determined uniquely by k and N . Naturally, as in any pattern classification task, several criteria for selecting these normalization factors may be considered and a deep study of their impact is out of the scope of this paper. However, it makes sense to

normalize $C_z(k)$ by dividing by (N, k) (N choose k), the number of terms summed in (2), which is also recursively computed: $(N, k) = (N - 1, k)N / (N - k)$. This way, we gain some robustness to deal with shapes described by different numbers of points.

The factor $1/N$ in the definition of shape power (3), which is not necessary in what respects to translation and scale invariance, also makes sense in what respects to dealing with different numbers of shape points (intuition comes by noting that replicating points will not change the defined power). Also, that factor is important to avoid the products in the elementary symmetric polynomials (2) to grow or decrease in prohibitively fast ways, due to points of very large or very small magnitude in the normalized shape (note that the mean of $\{|z_n|^2, 1 \leq n \leq N\}$ is unitary, regardless of $\{w_n\}$).

6. EXPERIMENTS

Since the subject of the paper is the generality of the properties of the proposed representation, we use distinct scenarios to illustrate the derived properties, rather than singling out a particular task to compare the performance against possible competitors. In particular, we report experiments that illustrate how nearest neighbor shape classification behaves in the presence of noise, the automatic clustering of binary images, and the capability of dealing with shapes that are extracted from real images with simple edge detection.

In what respects to the noise, we used the four shapes in the top row of the left side of Fig. 4 and performed 5000 tests that consisted in classifying randomly disturbed (*i.e.*, translated, rotated, and scaled) noisy versions of them, by simply measuring dissimilarity as the Frobenius norm of the difference of representations. The plot in Fig. 4 shows the percentage of correct classifications as a function of the noise level, showing 100% correct retrievals with noise standard deviation up to $\sigma = 0.25$, which is high enough to produce the perceptually misleading shapes in the bottom row of the left side of Fig. 4 (with the same size and orientation of the ones in the top).

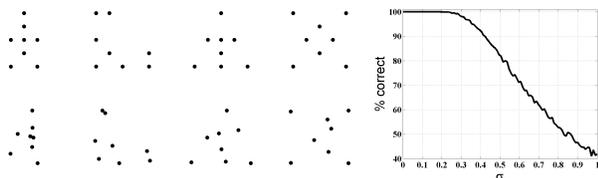


Fig. 4. Left: noise-free shapes (top) and noisy versions of them (bottom), with $\sigma = 0.25$, the approximate limit for 100% correct classifications. Right: accuracy as a function of the noise level.

The task just described is based on the comparison of pairs of shapes, thus it could be alternatively approached by attempting to compute the transformation between them (a non-trivial problem, as referred in Section 1). In opposition, most algorithms for tasks such as clustering, demand data represented in a way that factors out relevant transformations, so that statistics such as means, variances, etc. can be computed. To illustrate that our representation is adequate for this kind of tasks, we use 30×30 binary images obtained by thresholding greylevel images of digits with random orientations. The shapes extracted from these images are simply the sets of points corresponding to image pixels of value 1, which are not exactly related by a geometric transformation, due to the coarse discretization and binarization. Fig. 5 shows the result of a standard method used to automatically cluster the representations of all images. Note that the images corresponding to digits “6” and “9” are grouped in the same

cluster, which is not surprising, since they only differ by distinct orientations of the same pattern, thus having similar representations.

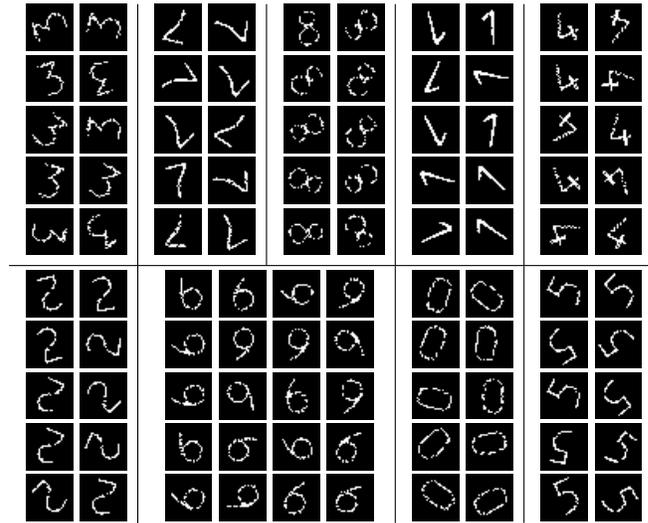


Fig. 5. Automatic clustering of binary images.

Finally, we describe an experiment where the shapes to classify are the edges of real images. Basically, we used (handheld) webcam images of trademark logos. Besides the distinct positions, sizes, and orientations of the logos, other disturbances come from the only approximate perpendicularity of the camera axis to the paper plane, which originates geometrically distorted shapes, and the sensitivity of the edge detection to illumination, resolution, etc. In spite of these disturbances, we were able to successfully classify several of the images by simply comparing the representations or the shapes corresponding to their edge points, as the examples in Fig. 6 illustrate.



Fig. 6. Examples of webcam images of logos correctly recognized.

7. CONCLUSION

We proposed a theoretically sound new representation scheme for shapes described by arbitrary collections of points, which is complete and invariant to arbitrary re-orderings of the shape points and geometric transformations such as translation, rotation, and scale.

8. REFERENCES

- [1] I. Bartolini, P. Ciaccia, and M. Patella, "Warp: Accurate Retrieval of Shapes using Phase of Fourier Descriptors and Time Warping Distance," *IEEE T-PAMI*, vol. 27, no. 1, 2005.
- [2] D. Kendall, D. Barden, T. Carne, and H. Le, *Shape and Shape Theory*, John Wiley and Sons, 1999.
- [3] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," *IEEE T-PAMI*, vol. 14, no. 2, 1992.
- [4] G. McNeill and S. Vijayakumar, "Hierarchical Procrustes Matching for Shape Retrieval," in *IEEE CVPR*, 2006.
- [5] T. Jebara, "Images as Bags of Pixels," in *IEEE ICCV*, 2003.
- [6] J. Flusser, B. Zitova, and T. Suk, *Moments and Moment Invariants in Pattern Recognition*, John Wiley & Sons, 2009.
- [7] J. Crespo and P. Aguiar, "Revisiting Complex Moments for 2D Shape Representation and Image Normalization," *IEEE T-IP*, vol. 20, no. 10, 2011.
- [8] J. Rodrigues, P. Aguiar, and J. Xavier, "ANSIG – An Analytic Signature for Permutation-Invariant Two-Dimensional Shape Representation," in *IEEE CVPR*, Anchorage, AK, USA, 2008.
- [9] B. Sadler and G. Giannakis, "Shift- and Rotation-invariant Object Reconstruction using the Bispectrum," *JOSA-A*, 1992.
- [10] V. Chandran, B. Carswell, B. Boashash, and S. Elgar, "Pattern Recognition using Invariants Defined from Higher Order Spectra: 2-D Image Inputs," *IEEE T-IP*, vol. 6, no. 5, 1997.
- [11] R. Kakarala, *Triple Correlation on Groups*, Ph.D. thesis, University of California, Irvine, 1992.
- [12] R. Kondor, *Group theoretical methods in machine learning*, Ph.D. thesis, Columbia University, 2008.
- [13] A. Oppenheim, R. Schaffer, and J. Buck, *Discrete-Time Signal Processing*, Prentice Hall, 1999.
- [14] J. Yellott and G. Iverson, "Uniqueness Properties of Higher-order Autocorrelation Functions," *JOSA-A*, vol. 9, no. 3, 1992.
- [15] K. Kanatani, *Group Theoretical Methods in Image Understanding*, Springer Verlag, Berlin, Germany, 1990.